GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 24, 2003, 15:32:50 ; Search time 15 Seconds
(without alignments)
33.346 Million cell updates/sec

Title: US-09-846-346-1
Perfect score: 88
Sequence: 1 SSKITHRIHWESASLLR 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	88	100.0	1663	2	US-08-793-126-1	Sequence 1, Appli
2	88	100.0	1663	4	US-09-132-271-1	Sequence 1, Appli
3	88	100.0	1663	4	US-09-142-334-22	Sequence 22, Appl
4	38	43.2	313	4	US-09-124-758-4	Sequence 4, Appli
5	37	42.0	150	4	US-09-605-785-707	Sequence 707, App
6	37	42.0	267	4	US-09-449-218D-43	Sequence 43, Appl
7	37	42.0	272	2	US-08-887-997B-2	Sequence 2, Appli
8	37	42.0	700	4	US-09-413-814-68	Sequence 68, Appl
9	37	42.0	751	4	US-08-969-415-2	Sequence 2, Appli
10	37	42.0	943	2	US-08-808-982-7	Sequence 7, Appli
11	37	42.0	943	4	US-09-306-902A-7	Sequence 7, Appli
12	36	40.9	20	3	US-08-504-538A-1	Sequence 1, Appli
13	36	40.9	20	3	US-08-504-538A-18	Sequence 18, Appl
14	36	40.9	20	4	US-09-249-458A-1	Sequence 1, Appli
15	36	40.9	20	4	US-08-630-052-1	Sequence 1, Appli
16	36	40.9	20	4	US-08-630-052-18	Sequence 18, Appl
17	36	40.9	20	5	PCT-US95-09307-1	Sequence 1, Appli
18	36	40.9	20	5	PCT-US95-09307-18	Sequence 18, Appl
19	36	40.9	24	3	US-08-504-538A-6	Sequence 6, Appli
20	36	40.9	24	4	US-09-249-458A-6	Sequence 6, Appli
21	36	40.9	24	4	US-08-630-052-6	Sequence 6, Appli
22	36	40.9	24	5	PCT-US95-09307-6	Sequence 6, Appli
23	36	40.9	93	1	US-08-839-710-3	Sequence 3, Appli
24	36	40.9	93	2	US-09-066-262-3	Sequence 3, Appli
25	36	40.9	166	2	US-08-729-103-4	Sequence 4, Appli
26	36	40.9	217	4	US-09-291-170A-4	Sequence 4, Appli
27	36	40.9	217	4	US-09-724-884-4	Sequence 4, Appli
28	36	40.9	516	4	US-09-291-170A-1	Sequence 1, Appli
29	36	40.9	516	4	US-09-724-884-1	Sequence 1, Appli
30	36	40.9	625	4	US-08-959-004-10	Sequence 10, Appl
31	36	40.9	844	1	US-07-646-537B-2	Sequence 2, Appli
32	36	40.9	893	4	US-09-514-302-4	Sequence 4, Appli
33	36	40.9	937	3	US-09-005-180A-4	Sequence 4, Appli
34	36	40.9	1190	1	US-08-337-690A-2	Sequence 2, Appli
35	36	40.9	1190	4	US-09-048-887-2	Sequence 2, Appli
36	36	40.9	1938	4	US-09-514-302-2	Sequence 2, Appli
37	35.5	40.3	24	4	US-09-082-279B-1199	Sequence 1199, Ap
38	35.5	40.3	24	4	US-09-315-304B-1199	Sequence 1199, Ap
39	35	39.8	117	6	5514582-15	Patent No. 5514582
40	35	39.8	165	2	US-08-401-530A-7	Sequence 7, Appli
41	35	39.8	165	2	US-08-729-103-3	Sequence 3, Appli
42	35	39.8	165	2	US-08-709-662-7	Sequence 7, Appli
43	35	39.8	212	4	US-08-861-774E-22	Sequence 22, Appl
44	35	39.8	212	4	US-08-861-774E-34	Sequence 34, Appl
45	35	39.8	233	4	US-09-214-631-7	Sequence 7, Appli

ALIGNMENTS

```
RESULT 1
US-08-793-126-1
; Sequence 1, Application US/08793126
; Patent No. 5849297
;  GENERAL INFORMATION:
;    APPLICANT:  Harrison, Richard Alexander
;    APPLICANT:  Farries, Charles Timothy
;    TITLE OF INVENTION:  MODIFIED HUMAN C3 PROTEINS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HALE AND DORR LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  United States of America
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/793,126
;      FILING DATE:  07-FEB-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Baker, Hollie L.
;      REGISTRATION NUMBER:  31,321
;      REFERENCE/DOCKET NUMBER:  102286.377
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 526-6000
;      TELEFAX:  (617) 526-5000
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1663 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-793-126-1

  Query Match             100.0%;  Score 88;  DB 2;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-06;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSKITHRIHWESASLLR 17
           |||||||||||||||||
Db    1304 SSKITHRIHWESASLLR 1320
```

```
RESULT 2
US-09-132-271-1
; Sequence 1, Application US/09132271
; Patent No. 6221657
;  GENERAL INFORMATION:
;    APPLICANT:  Harrison, Richard Alexander
;    APPLICANT:  Farries, Charles Timothy
;    TITLE OF INVENTION:  MODIFIED HUMAN C3 PROTEINS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HALE AND DORR LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  United States of America
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/132,271
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/793,126
;      FILING DATE:  07-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Baker, Hollie L.
;      REGISTRATION NUMBER:  31,321
;      REFERENCE/DOCKET NUMBER:  102286.377
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 526-6000
;      TELEFAX:  (617) 526-5000
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1663 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-132-271-1

  Query Match             100.0%;  Score 88;  DB 4;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-06;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSKITHRIHWESASLLR 17
           |||||||||||||||||
Db    1304 SSKITHRIHWESASLLR 1320


RESULT 3
US-09-142-334-22
; Sequence 22, Application US/09142334
; Patent No. 6268485
; GENERAL INFORMATION:
;  APPLICANT: Farries, Timothy C.
;  APPLICANT: Harrison, Richard A.
;  TITLE OF INVENTION: Down-Regulation Resistant C3 Convertase
;  FILE REFERENCE: 4-30443/A/IMU/PCT
;  CURRENT APPLICATION NUMBER: US/09/142,334
;  CURRENT FILING DATE: 1999-04-15
;  EARLIER APPLICATION NUMBER: PCT/GB97/00603
;  EARLIER FILING DATE: 1997-03-04
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
;   LENGTH: 1663
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-142-334-22

  Query Match             100.0%;  Score 88;  DB 4;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-06;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSKITHRIHWESASLLR 17
           |||||||||||||||||
Db    1304 SSKITHRIHWESASLLR 1320


RESULT 4
US-09-124-758-4
; Sequence 4, Application US/09124758
; Patent No. 6146849
;  GENERAL INFORMATION:
;    APPLICANT:  Pierce, J. M.
;    APPLICANT:  Moreman, Kelley W.
;    APPLICANT:  Lee, Jin-Kyu
;    TITLE OF INVENTION:  Lectins and Coding Sequences
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  US
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/124,758
;      FILING DATE:  04-JUN-1998
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/048,507
;      FILING DATE:  04-JUN-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  40-97
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  313 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-124-758-4

  Query Match             43.2%;  Score 38;  DB 4;  Length 313;
  Best Local Similarity   66.7%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       9 HWESASLLR 17
           || ::||||
Db     160 HWRNSSLLR 168


RESULT 5
US-09-605-785-707
; Sequence 707, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  Jiang, Yuqui
```

```
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Day, Craig H.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Li, Samuel
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Hepler, William
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C16
;  CURRENT APPLICATION NUMBER: US/09/605,785
;  CURRENT FILING DATE:  2000-06-27
;  NUMBER OF SEQ ID NOS: 835
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 707
;   LENGTH: 150
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-605-785-707

  Query Match             42.0%;  Score 37;  DB 4;  Length 150;
  Best Local Similarity   38.5%;  Pred. No. 72;
  Matches    5;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESA 13
          :  : || || :|
Db    130 AQSLAHRRHWRNA 142


RESULT 6
US-09-449-218D-43
; Sequence 43, Application US/09449218D
; Patent No. 6395511
; GENERAL INFORMATION:
;  APPLICANT: Brunkow, Mary E.
;  APPLICANT:  Galas, David J.
;  APPLICANT:  Kovacevich, Brian
;  APPLICANT:  Mulligan, John T.
;  APPLICANT:  Paeper, Bryan W.
;  APPLICANT:  Van Ness, Jeffrey
;  APPLICANT:  Winkler, David G.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INCREASING
;  TITLE OF INVENTION:  BONE MINERALIZATION
;  FILE REFERENCE: 240083.508
;  CURRENT APPLICATION NUMBER: US/09/449,218D
;  CURRENT FILING DATE:  1999-11-24
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 43
;   LENGTH: 267
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-449-218D-43

  Query Match             42.0%;  Score 37;  DB 4;  Length 267;
  Best Local Similarity   50.0%;  Pred. No. 1.3e+02;
  Matches    4;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      5 THRIHWES 12
          :| :|||:
Db    154 SHEVHWET 161


RESULT 7
US-08-887-997B-2
; Sequence 2, Application US/08887997B
; Patent No. 5935852
;  GENERAL INFORMATION:
;    APPLICANT:  FOLLETTIE, MAXIMILLIAN
;    APPLICANT:  DeROBERTIS, EDWARD M.
;    TITLE OF INVENTION:  Mammalian Cerberus-Like Protein &
;    TITLE OF INVENTION:  Compositions
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/887,997B
;      FILING DATE:  03-JUL-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAZAR, STEVEN R
;      REGISTRATION NUMBER:  32,618
;      REFERENCE/DOCKET NUMBER:  GI 5290
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8260
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  272 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-887-997B-2

  Query Match             42.0%;  Score 37;  DB 2;  Length 272;
  Best Local Similarity   50.0%;  Pred. No. 1.3e+02;
  Matches    4;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      5 THRIHWES 12
          :| :|||:
Db    154 SHEVHWET 161


RESULT 8
US-09-413-814-68
; Sequence 68, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
;  APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
;  APPLICANT: Bristol-Myers Squibb, Co.
;  APPLICANT: Beyer, Stefan
;  APPLICANT: Bloecker, Helmut
;  APPLICANT: Brandt, Petra
;  APPLICANT: Cino, Paul M
;  APPLICANT: Dougherty, Brian A
;  APPLICANT: Goldberg, Steven L
;  APPLICANT: Hofle, Gerhard
;  APPLICANT: Mueller, Joachim
;  APPLICANT: Reichenbach, Hans
;  TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
;  TITLE OF INVENTION: heteropolyketide compounds
;  FILE REFERENCE: PCT/US 99/23535
;  CURRENT APPLICATION NUMBER: US/09/413,814
;  CURRENT FILING DATE: 1999-10-07
;  EARLIER APPLICATION NUMBER: DE 198 46 493.2
;  EARLIER FILING DATE: 1998-10-09
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 68
;   LENGTH: 700
;   TYPE: PRT
```

```
;   ORGANISM: Sorangium cellulosum
US-09-413-814-68

  Query Match            42.0%;  Score 37;  DB 4;  Length 700;
  Best Local Similarity  60.0%;  Pred. No. 3.6e+02;
  Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       7 RIHWESASLL 16
           |:||: | ||
Db     269 RLHWDWAQLL 278


RESULT 9
US-08-969-415-2
; Sequence 2, Application US/08969415
; Patent No. 6410303
;  GENERAL INFORMATION:
;    APPLICANT:  TAKANO, Hiroyuki
;    APPLICANT:  HINO, Akihiro
;    APPLICANT:  IYO, Chie
;    APPLICANT:  SUZUKI, Yasuo
;    APPLICANT:  NAKAJIMA, Ryoichi
;    TITLE OF INVENTION:  FROZEN DOUGH-RESISTANT, PRACTICAL
;    TITLE OF INVENTION:  BAKER'S YEAST
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK, P.L.L.C.
;      STREET:  419 7th Street N.W., Ste. 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/969,415
;      FILING DATE:  21-OCT-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 08-297886
;      FILING DATE:  23-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEIMARK, Sheridan
;      REGISTRATION NUMBER:  20,520
;      REFERENCE/DOCKET NUMBER:  TAKANO=9
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 628-5197
;      TELEFAX:  (202) 737-3528
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  751 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-969-415-2

  Query Match            42.0%;  Score 37;  DB 4;  Length 751;
  Best Local Similarity  50.0%;  Pred. No. 3.9e+02;
  Matches    7;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 SSKITHRIHWESAS 14
           ::|| ||  :|||:
Db     559 NTKIKHRTSYESAT 572


RESULT 10
US-08-808-982-7
; Sequence 7, Application US/08808982
; Patent No. 5939271
;  GENERAL INFORMATION:
;    APPLICANT:  Tessier-Lavigne, Marc
;    APPLICANT:  Leonardo, E. David
;    APPLICANT:  Hink, Lindsay
;    APPLICANT:  Masu, Masayuki
;    APPLICANT:  Kazuko, Keino-Masu
;    TITLE OF INVENTION:  Netrin Receptors
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 BUSH STREET, SUITE 3200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/808,982
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UC96-217
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 343-4341
;      TELEFAX:  (415) 343-4342
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  943 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-808-982-7

  Query Match            42.0%;  Score 37;  DB 2;  Length 943;
  Best Local Similarity  57.1%;  Pred. No. 4.9e+02;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 SKITHRIHWESASL 15
           |: || :|  ||||
Db     514 SRDTHFLHLRSASL 527


RESULT 11
US-09-306-902A-7
; Sequence 7, Application US/09306902A
; Patent No. 6277585
;   GENERAL INFORMATION:
;        APPLICANT: Tessier-Lavigne, Marc
;                   Leonardo, E. David
;                   Hink, Lindsay
;                   Masu, Masayuki
;                   Kazuko, Keino-Masu
;        TITLE OF INVENTION: Netrin Receptors
;        NUMBER OF SEQUENCES: 9
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
;             STREET: 268 BUSH STREET, SUITE 3200
;             CITY: SAN FRANCISCO
;             STATE: CALIFORNIA
;             COUNTRY: USA
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
```

```
;             APPLICATION NUMBER: US/09/306,902A
;             FILING DATE: 07-May-1999
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: OSMAN, RICHARD A
;             REGISTRATION NUMBER: 36,627
;             REFERENCE/DOCKET NUMBER: UC96-217
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 343-4341
;             TELEFAX: (415) 343-4342
;   INFORMATION FOR SEQ ID NO: 7:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 943 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: not relevant
;             TOPOLOGY: not relevant
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-306-902A-7

  Query Match             42.0%;  Score 37;  DB 4;  Length 943;
  Best Local Similarity   57.1%;  Pred. No. 4.9e+02;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 SKITHRIHWESASL 15
          |: || :|  ||||
Db    514 SRDTHFLHLRSASL 527


RESULT 12
US-08-504-538A-1
; Sequence 1, Application US/08504538A
; Patent No. 6004746
;  GENERAL INFORMATION:
;    APPLICANT:  Brent, Roger
;    APPLICANT:  McCoy, John M.
;    APPLICANT:  Jessen, Timm H.
;    TITLE OF INVENTION:  INTERACTION TRAP SYSTEMS FOR DETECTING
;    TITLE OF INVENTION:  PROTEIN INTERACTIONS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Clark & Elbing LLP
;      STREET:  176 Federal Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02110-2214
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/504,538A
;      FILING DATE:  07/20/95
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/278,082
;      FILING DATE:  07/20/94
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul T. Clark
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00786/259001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 428-0200
;      TELEFAX:  (617) 428-7045
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-504-538A-1

  Query Match             40.9%;  Score 36;  DB 3;  Length 20;
  Best Local Similarity   38.5%;  Pred. No. 13;
  Matches    5;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      5 THRIHWESASLLR 17
          ::|: ||: :| |
Db      5 SYRLDWEAGALFR 17


RESULT 13
US-08-504-538A-18
; Sequence 18, Application US/08504538A
; Patent No. 6004746
;  GENERAL INFORMATION:
;    APPLICANT:  Brent, Roger
;    APPLICANT:  McCoy, John M.
;    APPLICANT:  Jessen, Timm H.
;    TITLE OF INVENTION:  INTERACTION TRAP SYSTEMS FOR DETECTING
;    TITLE OF INVENTION:  PROTEIN INTERACTIONS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Clark & Elbing LLP
;      STREET:  176 Federal Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02110-2214
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/504,538A
;      FILING DATE:  07/20/95
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/278,082
;      FILING DATE:  07/20/94
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul T. Clark
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00786/259001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 428-0200
;      TELEFAX:  (617) 428-7045
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-504-538A-18

  Query Match             40.9%;  Score 36;  DB 3;  Length 20;
  Best Local Similarity   38.5%;  Pred. No. 13;
  Matches    5;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      5 THRIHWESASLLR 17
          ::|: ||: :| |
Db      5 SYRLDWEAGALFR 17


RESULT 14
US-09-249-458A-1
; Sequence 1, Application US/09249458A
; Patent No. 6242183
; GENERAL INFORMATION:
;  APPLICANT: Brent, Roger
;  APPLICANT: Jessen, Timm H.
```

```
;  APPLICANT: McCoy, John M.
;  TITLE OF INVENTION: INTERACTION TRAP SYSTEMS FOR DETECTING
;  TITLE OF INVENTION: PROTEIN INTERACTIONS
;  FILE REFERENCE: 00786/222002
;  CURRENT APPLICATION NUMBER: US/09/249,458A
;  CURRENT FILING DATE: 1999-02-12
;  EARLIER APPLICATION NUMBER: 08/278,082
;  EARLIER FILING DATE: 1994-07-20
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 20
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-249-458A-1

  Query Match             40.9%;  Score 36;  DB 4;  Length 20;
  Best Local Similarity   38.5%;  Pred. No. 13;
  Matches    5;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 THRIHWESASLLR 17
           ::|: ||: :| |
Db       5 SYRLDWEAGALFR 17


RESULT 15
US-08-630-052-1
; Sequence 1, Application US/08630052
; Patent No. 6399296
;  GENERAL INFORMATION:
;    APPLICANT:  Brent, Roger
;    APPLICANT:  McCoy, John M.
;    APPLICANT:  Jessen, Timm H.
;    APPLICANT:  Xu, Chanxing Wilson
;    TITLE OF INVENTION:  INTERACTION TRAP SYSTEMS FOR DETECTING PROTEIN
;    TITLE OF INVENTION:  INTERACTIONS
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/630,052
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/504,538
;      FILING DATE:  July 20, 1995
;      APPLICATION NUMBER:  08/278,082
;      FILING DATE:  July 20, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Karen F. Lech
;      REGISTRATION NUMBER:  35,238
;      REFERENCE/DOCKET NUMBER:  00786/311001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
US-08-630-052-1

  Query Match             40.9%;  Score 36;  DB 4;  Length 20;
  Best Local Similarity   38.5%;  Pred. No. 13;
  Matches    5;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 THRIHWESASLLR 17
           ::|: ||: :| |
Db       5 SYRLDWEAGALFR 17


Search completed: February 24, 2003, 15:35:06
Job time : 16 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 24, 2003, 15:33:10 ; Search time 12 Seconds
(without alignments)
44.016 Million cell updates/sec

Title: US-09-846-346-1
Perfect score: 88
Sequence: 1 SSKITHRIHWESASLLR 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 156504 seqs, 31069816 residues

Total number of hits satisfying chosen parameters: 156504

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	88	100.0	17	9	US-09-846-346-1	Sequence 1, Appli
2	88	100.0	1663	10	US-09-875-519A-22	Sequence 22, Appl
3	83	94.3	16	9	US-09-846-345-1	Sequence 1, Appli
4	75	85.2	14	9	US-09-845-730-1	Sequence 1, Appli
5	66	75.0	12	9	US-09-846-349-1	Sequence 1, Appli
6	61	69.3	11	9	US-09-845-715-1	Sequence 1, Appli
7	53	60.2	10	9	US-09-845-731-1	Sequence 1, Appli
8	42	47.7	91	10	US-09-867-550-910	Sequence 910, App
9	41	46.6	74	10	US-09-764-864-816	Sequence 816, App
10	40	45.5	64	10	US-09-867-550-908	Sequence 908, App
11	39	44.3	2012	9	US-09-808-602-68	Sequence 68, Appl
12	38	43.2	34	10	US-09-864-761-39376	Sequence 39376, A
13	38	43.2	66	9	US-09-796-692-1309	Sequence 1309, Ap
14	38	43.2	66	9	US-09-796-692-1782	Sequence 1782, Ap
15	38	43.2	66	9	US-09-796-692-2050	Sequence 2050, Ap
16	38	43.2	66	9	US-09-796-692-2300	Sequence 2300, Ap
17	38	43.2	313	9	US-09-992-598-414	Sequence 414, App
18	38	43.2	313	9	US-09-989-293A-414	Sequence 414, App
19	38	43.2	313	9	US-10-063-547-88	Sequence 88, Appl
20	38	43.2	313	9	US-09-989-735-414	Sequence 414, App
21	38	43.2	313	9	US-09-990-444-414	Sequence 414, App
22	38	43.2	313	9	US-09-989-730-414	Sequence 414, App
23	38	43.2	313	9	US-09-990-436-414	Sequence 414, App
24	38	43.2	313	9	US-09-991-181-414	Sequence 414, App
25	38	43.2	313	9	US-09-993-687-414	Sequence 414, App
26	38	43.2	313	9	US-09-989-734-414	Sequence 414, App
27	38	43.2	313	9	US-09-997-653-414	Sequence 414, App
28	38	43.2	313	9	US-10-174-590-294	Sequence 294, App
29	38	43.2	313	9	US-10-176-758-294	Sequence 294, App
30	38	43.2	313	9	US-10-063-616-88	Sequence 88, Appl
31	38	43.2	313	9	US-10-175-737-294	Sequence 294, App
32	38	43.2	313	9	US-09-993-667-414	Sequence 414, App
33	38	43.2	313	9	US-10-063-502-88	Sequence 88, Appl
34	38	43.2	313	9	US-10-173-706-294	Sequence 294, App
35	38	43.2	313	9	US-10-175-738-294	Sequence 294, App
36	38	43.2	313	9	US-10-175-752-294	Sequence 294, App
37	38	43.2	313	9	US-10-176-482-294	Sequence 294, App
38	38	43.2	313	9	US-10-176-757-294	Sequence 294, App
39	38	43.2	313	9	US-10-176-913-294	Sequence 294, App
40	38	43.2	313	9	US-10-180-552-294	Sequence 294, App
41	38	43.2	313	9	US-10-180-557-294	Sequence 294, App
42	38	43.2	313	9	US-09-990-438-414	Sequence 414, App
43	38	43.2	313	9	US-09-990-562-414	Sequence 414, App
44	38	43.2	313	9	US-09-997-428-414	Sequence 414, App
45	38	43.2	313	9	US-09-997-666-414	Sequence 414, App

ALIGNMENTS

```
RESULT 1
US-09-846-346-1
; Sequence 1, Application US/09846346
; Patent No. US20020160532A1
; GENERAL INFORMATION:
;  APPLICANT: Jackowski, George
;  TITLE OF INVENTION: BIOPOLYMER MARKER INDICATIVE OF DISEASE STATE HAVING A MOLECUL
;  TITLE OF INVENTION:  OF 1998 DALTONS
;  FILE REFERENCE: 2132.013
;  CURRENT APPLICATION NUMBER: US/09/846,346
;  CURRENT FILING DATE:  2001-04-30
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 17
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-846-346-1

  Query Match             100.0%;  Score 88;  DB 9;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-08;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSKITHRIHWESASLLR 17
           |||||||||||||||||
Db       1 SSKITHRIHWESASLLR 17


RESULT 2
US-09-875-519A-22
; Sequence 22, Application US/09875519A
; Patent No. US20020068059A1
; GENERAL INFORMATION:
;  APPLICANT: Farries, Timothy C.
;  APPLICANT:  Harrison, Richard A.
;  TITLE OF INVENTION: Down-Regulation Resistant C3 Convertase
;  FILE REFERENCE: 4-30443/A/IMU/PCT
;  CURRENT APPLICATION NUMBER: US/09/875,519A
;  CURRENT FILING DATE:  2001-06-06
;  PRIOR APPLICATION NUMBER: PCT/GB97/00603
;  PRIOR FILING DATE: 1997-03-04
;  NUMBER OF SEQ ID NOS: 35
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
;   LENGTH: 1663
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-875-519A-22

  Query Match             100.0%;  Score 88;  DB 10;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 7.5e-06;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESASLLR 17
          |||||||||||||||||
Db   1304 SSKITHRIHWESASLLR 1320


RESULT 3
US-09-846-345-1
; Sequence 1, Application US/09846345
; Patent No. US20020161182A1
; GENERAL INFORMATION:
;  APPLICANT: Jackowski, George
;  TITLE OF INVENTION: BIOPOLYMER MARKER INDICATIVE OF DISEASE STATE HAVING A MOLECULAR
;  TITLE OF INVENTION:  OF 1865 DALTONS
;  FILE REFERENCE: 2132.045
;  CURRENT APPLICATION NUMBER: US/09/846,345
;  CURRENT FILING DATE:  2001-04-30
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 16
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-846-345-1

  Query Match             94.3%;  Score 83;  DB 9;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESASLL 16
          ||||||||||||||||
Db      1 SSKITHRIHWESASLL 16


RESULT 4
US-09-845-730-1
; Sequence 1, Application US/09845730
; Patent No. US20020169278A1
; GENERAL INFORMATION:
;  APPLICANT: Jackowski, George
;  TITLE OF INVENTION: BIOPOLYMER MARKER INDICATIVE OF DISEASE STATE HAVING A MOLECULAR
;  TITLE OF INVENTION:  OF 1690 DALTONS
;  FILE REFERENCE: 2132.042
;  CURRENT APPLICATION NUMBER: US/09/845,730
;  CURRENT FILING DATE:  2001-04-30
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 14
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-845-730-1

  Query Match             85.2%;  Score 75;  DB 9;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 6.7e-06;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 KITHRIHWESASLL 16
          ||||||||||||||
Db      1 KITHRIHWESASLL 14


RESULT 5
```

```
US-09-846-349-1
; Sequence 1, Application US/09846349
; Patent No. US20020161186A1
; GENERAL INFORMATION:
;  APPLICANT: Jackowski, George
;  TITLE OF INVENTION: BIOPOLYMER MARKER INDICATIVE OF DISEASE STATE HAVING A MOLECUL
;  TITLE OF INVENTION:  OF 1449 DALTONS
;  FILE REFERENCE: 2132.034
;  CURRENT APPLICATION NUMBER: US/09/846,349
;  CURRENT FILING DATE:  2001-04-30
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 12
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-846-349-1

  Query Match             75.0%;  Score 66;  DB 9;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 0.00014;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 THRIHWESASLL 16
          ||||||||||||
Db      1 THRIHWESASLL 12


RESULT 6
US-09-845-715-1
; Sequence 1, Application US/09845715
; Patent No. US20020161184A1
; GENERAL INFORMATION:
;  APPLICANT: Jackowski, George
;  TITLE OF INVENTION: BIOPOLYMER MARKER INDICATIVE OF DISEASE STATE HAVING A MOLECUL
;  TITLE OF INVENTION:  OF 1348 DALTONS
;  FILE REFERENCE: 2132.030
;  CURRENT APPLICATION NUMBER: US/09/845,715
;  CURRENT FILING DATE:  2001-04-30
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 11
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-845-715-1

  Query Match             69.3%;  Score 61;  DB 9;  Length 11;
  Best Local Similarity   100.0%;  Pred. No. 0.00079;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 HRIHWESASLL 16
          ||||||||||||
Db      1 HRIHWESASLL 11


RESULT 7
US-09-845-731-1
; Sequence 1, Application US/09845731
; Publication No. US20030004307A1
; GENERAL INFORMATION:
;  APPLICANT: Jackowski, George
;  TITLE OF INVENTION: Biopolymer Marker Indicative Of Disease State Having A Molecul
;  TITLE OF INVENTION:  Of 1211 Daltons
;  FILE REFERENCE: 2132.029
;  CURRENT APPLICATION NUMBER: US/09/845,731
;  CURRENT FILING DATE:  2001-04-30
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 10
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-845-731-1
```

```
  Query Match             60.2%;  Score 53;  DB 9;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 RIHWESASLL 16
          ||||||||||
Db      1 RIHWESASLL 10


RESULT 8
US-09-867-550-910
; Sequence 910, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
;  APPLICANT: Leach, Martin D.
;  APPLICANT:  Mehraban, Fuad,
;  APPLICANT:  Conley, Pamela
;  APPLICANT:  Law, Debbie
;  APPLICANT:  Topper, James
;  TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells and
;  TITLE OF INVENTION:  Thereby
;  FILE REFERENCE: 21402-013 (Cura-313)
;  CURRENT APPLICATION NUMBER: US/09/867,550
;  CURRENT FILING DATE:  2001-09-20
;  PRIOR APPLICATION NUMBER: USSN 60/208,427
;  PRIOR FILING DATE: 2000-05-30
;  NUMBER OF SEQ ID NOS: 2125
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 910
;   LENGTH: 91
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)
;   OTHER INFORMATION: wherein Xaa may be any one of Arg or Cys or Gly or Ser
US-09-867-550-910

  Query Match             47.7%;  Score 42;  DB 10;  Length 91;
  Best Local Similarity   50.0%;  Pred. No. 6;
  Matches    7;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      2 SKITHRIHWESASL 15
          ||:  | ||:|  |
Db     72 SKVCSRFHWDSGVL 85


RESULT 9
US-09-764-864-816
; Sequence 816, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 816
;   LENGTH: 74
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (23)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-816

  Query Match             46.6%;  Score 41;  DB 10;  Length 74;
  Best Local Similarity   53.3%;  Pred. No. 6.9;
  Matches    8;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESASL 15
          ::|| | :| :||||
Db     35 ATKIKHFLHQQSASL 49


RESULT 10
US-09-867-550-908
; Sequence 908, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
;  APPLICANT: Leach, Martin D.
;  APPLICANT:  Mehraban, Fuad,
;  APPLICANT:  Conley, Pamela
;  APPLICANT:  Law, Debbie
;  APPLICANT:  Topper, James
;  TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells a
;  TITLE OF INVENTION:  Thereby
;  FILE REFERENCE: 21402-013 (Cura-313)
;  CURRENT APPLICATION NUMBER: US/09/867,550
;  CURRENT FILING DATE:  2001-09-20
;  PRIOR APPLICATION NUMBER: USSN 60/208,427
;  PRIOR FILING DATE: 2000-05-30
;  NUMBER OF SEQ ID NOS: 2125
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 908
;   LENGTH: 64
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-867-550-908

  Query Match             45.5%;  Score 40;  DB 10;  Length 64;
  Best Local Similarity   43.8%;  Pred. No. 8.6;
  Matches    7;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 SKITHRIHWESASLLR 17
          :::|||:|  ||:  |
Db     19 NRLTHRVHSASANAPR 34


RESULT 11
US-09-808-602-68
; Sequence 68, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
;  APPLICANT: Vernet, Corine A
;  APPLICANT:  Fernandes, Elma
;  APPLICANT:  Shimkets, Richard A
;  APPLICANT:  Herrman, John L
;  APPLICANT:  Majumder, Kumud
;  APPLICANT:  Mishra, Vishnu
;  APPLICANT:  Mezes, Peter S
;  APPLICANT:  MacDougall, John
;  TITLE OF INVENTION: No. US20020155115A1el Proteins and Nuclec Acids Encoding Same
;  FILE REFERENCE: 15966-697 CIP
;  CURRENT APPLICATION NUMBER: US/09/808,602
;  CURRENT FILING DATE:  2001-03-14
;  PRIOR APPLICATION NUMBER: 09/800,198
;  PRIOR FILING DATE: 2001-03-05
;  PRIOR APPLICATION NUMBER: 60/186,596
;  PRIOR FILING DATE: 2000-03-03
;  NUMBER OF SEQ ID NOS: 114
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 68
;   LENGTH: 2012
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-808-602-68

  Query Match             44.3%;  Score 39;  DB 9;  Length 2012;
  Best Local Similarity   42.9%;  Pred. No. 3.9e+02;
  Matches    6;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy      1 SSKITHRIHWESAS 14
          |  :|| :|::| |
Db   1699 SLTVTHTVHYQSVS 1712


RESULT 12
US-09-864-761-39376
; Sequence 39376, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39376
;   LENGTH: 34
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC004098.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.7
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.1
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.7
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.5
;   OTHER INFORMATION: EST_HUMAN HIT: AW615804.1, EVALUE 2.00e-13
;   OTHER INFORMATION: SWISSPROT HIT: O93477, EVALUE 3.80e+00
US-09-864-761-39376

  Query Match             43.2%;  Score 38;  DB 10;  Length 34;
  Best Local Similarity   35.3%;  Pred. No. 9.3;
  Matches    6;  Conservative    5;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESASLLR 17
          |::|  :: ||:  | |
Db     12 STRIGEKVQWEACRLYR 28


RESULT 13
US-09-796-692-1309
; Sequence 1309, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
;  APPLICANT: Gaiger, Alexander
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
;  TITLE OF INVENTION:  HEMATOLOGICAL MALIGNANCIES
;  FILE REFERENCE: 2077.001200
;  CURRENT APPLICATION NUMBER: US/09/796,692
;  CURRENT FILING DATE:  2001-03-01
;  PRIOR APPLICATION NUMBER: 60/186,126
;  PRIOR FILING DATE: 2000-03-01
;  PRIOR APPLICATION NUMBER: 60/190,479
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: 60/200,545
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 60/200,303
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: 60/200,779
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: 60/200,999
;  PRIOR FILING DATE: 2000-05-01
;  PRIOR APPLICATION NUMBER: 60/202,084
;  PRIOR FILING DATE: 2000-05-04
;  PRIOR APPLICATION NUMBER: 60/206,201
;  PRIOR FILING DATE: 2000-05-22
;  PRIOR APPLICATION NUMBER: 60/218,950
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: 60/222,903
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: 60/223,416
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: 60/223,378
;  PRIOR FILING DATE: 2000-08-07
;  NUMBER OF SEQ ID NOS: 9597
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1309
;   LENGTH: 66
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-796-692-1309

  Query Match             43.2%;  Score 38;  DB 9;  Length 66;
  Best Local Similarity   50.0%;  Pred. No. 18;
  Matches    7;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESAS 14
          |:|:||   | :||
Db     42 SAKLTHCTTWAAAS 55


RESULT 14
US-09-796-692-1782
; Sequence 1782, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
;  APPLICANT: Gaiger, Alexander
;  APPLICANT:  Algate, Paul A.
```

```
;  APPLICANT:  Mannion, Jane
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
;  TITLE OF INVENTION:  HEMATOLOGICAL MALIGNANCIES
;  FILE REFERENCE: 2077.001200
;  CURRENT APPLICATION NUMBER: US/09/796,692
;  CURRENT FILING DATE:  2001-03-01
;  PRIOR APPLICATION NUMBER: 60/186,126
;  PRIOR FILING DATE: 2000-03-01
;  PRIOR APPLICATION NUMBER: 60/190,479
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: 60/200,545
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 60/200,303
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: 60/200,779
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: 60/200,999
;  PRIOR FILING DATE: 2000-05-01
;  PRIOR APPLICATION NUMBER: 60/202,084
;  PRIOR FILING DATE: 2000-05-04
;  PRIOR APPLICATION NUMBER: 60/206,201
;  PRIOR FILING DATE: 2000-05-22
;  PRIOR APPLICATION NUMBER: 60/218,950
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: 60/222,903
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: 60/223,416
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: 60/223,378
;  PRIOR FILING DATE: 2000-08-07
;  NUMBER OF SEQ ID NOS: 9597
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1782
;   LENGTH: 66
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-796-692-1782

  Query Match             43.2%;  Score 38;  DB 9;  Length 66;
  Best Local Similarity   50.0%;  Pred. No. 18;
  Matches    7;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 SSKITHRIHWESAS 14
           |:|:||   | :||
Db      42 SAKLTHCTTWAAAS 55


RESULT 15
US-09-796-692-2050
; Sequence 2050, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
;  APPLICANT: Gaiger, Alexander
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
;  TITLE OF INVENTION:  HEMATOLOGICAL MALIGNANCIES
;  FILE REFERENCE: 2077.001200
;  CURRENT APPLICATION NUMBER: US/09/796,692
;  CURRENT FILING DATE:  2001-03-01
;  PRIOR APPLICATION NUMBER: 60/186,126
;  PRIOR FILING DATE: 2000-03-01
;  PRIOR APPLICATION NUMBER: 60/190,479
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: 60/200,545
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 60/200,303
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: 60/200,779
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: 60/200,999
;  PRIOR FILING DATE: 2000-05-01
;  PRIOR APPLICATION NUMBER: 60/202,084
;  PRIOR FILING DATE: 2000-05-04
;  PRIOR APPLICATION NUMBER: 60/206,201
;  PRIOR FILING DATE: 2000-05-22
;  PRIOR APPLICATION NUMBER: 60/218,950
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: 60/222,903
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: 60/223,416
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: 60/223,378
;  PRIOR FILING DATE: 2000-08-07
;  NUMBER OF SEQ ID NOS: 9597
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2050
;   LENGTH: 66
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-796-692-2050

  Query Match             43.2%;  Score 38;  DB 9;  Length 66;
  Best Local Similarity   50.0%;  Pred. No. 18;
  Matches    7;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 SSKITHRIHWESAS 14
           |:|:||   | :||
Db      42 SAKLTHCTTWAAAS 55


Search completed: February 24, 2003, 15:35:24
Job time : 13 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 24, 2003, 15:32:50 ; Search time 15 Seconds
(without alignments)
108.952 Million cell updates/sec

Title: US-09-846-346-1
Perfect score: 88
Sequence: 1 SSKITHRIHWESASLLR 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	88	100.0	1663	1	C3HU	complement C3 prec
2	61	69.3	726	2	A27602	complement C3 - ra
3	52	59.1	267	2	A82997	hypothetical prote
4	46	52.3	1663	1	C3RT	complement C3 prec
5	45	51.1	211	2	H83239	pseudouridine synt
6	45	51.1	336	2	F75508	mrr restriction sy
7	45	51.1	1663	1	C3MS	complement C3 prec
8	44	50.0	516	2	S67037	SMP3 protein - yea
9	42	47.7	248	2	AH0011	ferredoxin-NADP re
10	42	47.7	280	2	C86317	protein T10O22.23
11	42	47.7	401	2	E82521	hypothetical prote
12	42	47.7	474	2	G75580	conserved hypothet
13	42	47.7	858	2	T18946	probable phospholi
14	41	46.6	226	1	JQ0393	nodulation protein
15	41	46.6	229	2	AI3289	hypothetical cytos
16	41	46.6	615	2	B86713	hypothetical prote
17	41	46.6	1585	2	H97690	NAD-glutamate dehy
18	41	46.6	1585	2	AE2916	NAD-glutamate dehy
19	41	46.6	1666	1	C3GP	complement C3 prec
20	40.5	46.0	1417	2	H90670	probable invasin [
21	40.5	46.0	1417	2	D85521	probable adhesin e
22	40	45.5	259	2	T29569	hypothetical prote
23	40	45.5	343	2	T42129	probable acyltrans
24	40	45.5	354	2	D41080	probable aldolase
25	40	45.5	593	2	C97848	ABC transporter AT
26	40	45.5	1123	2	T22608	hypothetical prote
27	40	45.5	1456	2	G86466	hypothetical prote
28	40	45.5	2514	2	T37320	ataxia telangiecta
29	40	45.5	2619	2	T24588	hypothetical prote
30	39	44.3	228	2	AI2913	conserved hypothet
31	39	44.3	249	2	T16228	hypothetical prote
32	39	44.3	263	2	T48742	hypothetical prote
33	39	44.3	266	2	D97688	hypothetical prote
34	39	44.3	406	2	T50894	hydroxyneurosporen
35	39	44.3	459	2	B82416	hypothetical prote
36	39	44.3	493	2	G88979	protein F37B4.5 [i
37	39	44.3	567	2	C69611	ABC transporter re
38	39	44.3	574	2	AB1790	ABC transporter re
39	39	44.3	574	2	AC1414	ABC transporter re
40	39	44.3	790	2	S18206	recombination prot
41	39	44.3	851	1	WMBEU9	gene UL9 protein -
42	39	44.3	1015	1	TQECT	transposase - Esch
43	39	44.3	1479	2	T17401	transcription regu
44	39	44.3	1896	2	T08851	Down syndrome cell
45	38	43.2	148	2	A86878	non-heme iron-bind

ALIGNMENTS

RESULT 1
C3HU
complement C3 precursor [validated] - human
N;Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) C3b subuni
C;Species: Homo sapiens (man)
C;Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 08-Dec-2000
C;Accession: A94065; A37999; A92187; A27603; A23435; A45830; B45830; A01257; A01258
R;de Bruijn, M.H.L.; Fey, G.H.
Proc. Natl. Acad. Sci. U.S.A. 82, 708-712, 1985
A;Title: Human complement component C3: cDNA coding sequence and derived primary stru
A;Reference number: A94065; MUID:85140166; PMID:2579379
A;Accession: A94065
A;Molecule type: mRNA
A;Residues: 1-1663 <DEB>
A;Cross-references: GB:K02765; NID:g179664; PIDN:AAA85332.1; PID:g179665
R;Vik, D.P.; Amiguet, P.; Moffat, G.J.; Fey, M.; Amiguet-Barras, F.; Wetsel, R.A.; Ta
Biochemistry 30, 1080-1085, 1991
A;Title: Structural features of the human C3 gene: intron/exon organization, transcri
A;Reference number: A37999; MUID:91113687; PMID:1703437
A;Contents: intron/exon structure of gene
A;Accession: A37999
A;Molecule type: DNA
A;Residues: 1-25 <VIK>
A;Cross-references: GB:M63423
A;Note: the authors translated the codon GGT for residue 6 as Leu, CCC for residue 7
R;Hugli, T.E.
J. Biol. Chem. 250, 8293-8301, 1975
A;Title: Human anaphylatoxin (C3a) from the third component of complement.
A;Reference number: A92187; MUID:76069169; PMID:1238393
A;Accession: A92187
A;Molecule type: protein
A;Residues: 672-680,'N',682-699,'Q',701-748 <HUG>
R;Daoudaki, M.E.; Becherer, J.D.; Lambris, J.D.
J. Immunol. 140, 1577-1580, 1988
A;Title: A 34-amino acid peptide of the third component of complement mediates proper
A;Reference number: A27603; MUID:88154452; PMID:3279119
A;Accession: A27603
A;Molecule type: protein
A;Residues: 1409-1563 <DAO>
R;Hellman, U.; Eggertsen, G.; Engstrom, A.; Sjoquist, J.
Biochem. J. 230, 353-361, 1985
A;Title: Amino acid sequence of the trypsin-generated C3d fragment from human comple
A;Reference number: A23435; MUID:86025442; PMID:3876831
A;Accession: A23435
A;Molecule type: protein
A;Residues: 1002-1012,'E',1014-1303 <HEL>
A;Note: sequence corresponding to residues 1072-1100 was not determined but was take
R;Poznansky, M.C.; Clissold, P.M.; Lachmann, P.J.
J. Immunol. 143, 1254-1258, 1989
A;Title: The difference between human C3F and C3S results from a single amino acid c
3.
A;Reference number: A45830; MUID:89309808; PMID:2473125

A;Accession: A45830
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1212-1215,'N',1217-1223 <POZ>
A;Note: this is the C3S allele
A;Accession: B45830
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1212-1223 <PO2>
R;Dolmer, K.; Sottrup-Jensen, L.
FEBS Lett. 315, 85-90, 1993
A;Title: Disulfide bridges in human complement component C3b.
A;Reference number: S27041; MUID:93106233; PMID:8416818
A;Contents: annotation; disulfide bonds
C;Comment: The sequence shown is the C3 fast (C3F) allele, which is found mainly in Cauc
C;Comment: Complement C3 contains two chains, formed by removal of four residues and lin
alternative complement pathways, releases the C3a anaphylatoxin from the amino end of t
rnative-complement-pathway C3/C5 convertase.
C;Comment: C3a anaphylatoxin is a vasoactive peptide and a mediator of inflammation.
C;Comment: C3b, with its highly reactive thiol group, binds to the surface of foreign pa
e classical-complement-pathway C3/C5 convertase. The activity of C3b is regulated by prc
C;Comment: The major site of synthesis of this plasma protein is the liver.
C;Genetics:
A;Gene: GDB:C3
A;Cross-references: GDB:119044; OMIM:120700
A;Map position: 19p13.3-19p13.3
A;Note: contains 41 exons
C;Superfamily: alpha-2-macroglobulin
C;Keywords: acute phase; complement alternate pathway; complement pathway; glycoprotein;
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-667/Product: complement C3 and C3b beta chain #status predicted <C3BB>
F;23-667,672-1663/Product: complement C3 #status predicted <CC3>
F;23-667,749-1663/Product: C3b #status predicted <C3B>
F;672-1663/Product: complement C3 alpha chain #status predicted <CC3A>
F;672-748/Product: C3a anaphylatoxin #status predicted <C3T>
F;749-1663/Product: C3b alpha' chain #status predicted <C3BA>
F;946-1303/Product: C3dk fragment #status predicted <CDK>
F;955-1303/Product: C3dg fragment #status predicted <CDG>
F;955-1001/Product: C3g fragment #status predicted <C3G>
F;1002-1303/Product: C3d fragment #status experimental <C3D>
F;1424-1457/Region: properdin binding
F;85,939/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;559-816,627-662,693-720,694-727,707-728,873-1513,1101-1158,1358-1489,1389-1458,1506-15
F;748-749/Cleavage site: Arg-Ser (C3 convertase) #status predicted
F;954-955/Cleavage site: Arg-Glu (complement factor I) #status predicted
F;1010-1013/Cross-link: thiolester (Cys-Gln) #status experimental
F;1303-1304/Cleavage site: Arg-Ser (complement factor I) #status predicted
F;1320-1321/Cleavage site: Arg-Ser (complement factor I) #status predicted
F;1617/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match 100.0%; Score 88; DB 1; Length 1663;
  Best Local Similarity 100.0%; Pred. No. 1e-06;
  Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSKITHRIHWESASLLR 17
||||||||||||||||||
Db 1304 SSKITHRIHWESASLLR 1320

RESULT 2
A27602
complement C3 - rabbit (fragment)
N;Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) C3b subunit;
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 15-Dec-1988 #sequence_revision 07-Oct-1994 #text_change 16-Jul-1999
C;Accession: A27602
R;Kusano, M.; Choi, N.H.; Tomita, M.; Yamamoto, K.; Migita, S.; Sekiya, T.; Nishimura, S
Immunol. Invest. 15, 365-378, 1986
A;Title: Nucleotide sequence of cDNA and derived amino acid sequence of rabbit complemen
A;Reference number: A27602; MUID:87006907; PMID:3019881
A;Accession: A27602
A;Molecule type: mRNA
A;Residues: 1-726 <KUS>
A;Cross-references: GB:M32434; NID:g164862; PIDN:AAA31190.1; PID:g164863
C;Comment: Complement C3 contains two chains, formed by removal of four residues and
alternative complement pathways, releases the C3a anaphylatoxin from the amino end o
rnative-complement-pathway C3/C5 convertase.
C;Comment: C3a anaphylatoxin is a vasoactive peptide and a mediator of inflammation.
C;Comment: C3b, with its highly reactive thiol group, binds to the surface of foreign
e classical-complement-pathway C3/C5 convertase. The activity of C3b is regulated by
C;Comment: The major site of synthesis of this plasma protein is the liver.
C;Superfamily: alpha-2-macroglobulin
C;Keywords: acute phase; complement alternate pathway; complement pathway; glycoprote

  Query Match 69.3%; Score 61; DB 2; Length 726;
  Best Local Similarity 70.6%; Pred. No. 0.019;
  Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 SSKITHRIHWESASLLR 17
|| : ||| |:||||||
Db 367 SSPVKHRIVWDSASLLR 383

RESULT 3
A82997
hypothetical protein PA5194 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A82997
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: A82997
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-267 <STO>
A;Cross-references: GB:AE004932; GB:AE004091; NID:g9951493; PIDN:AAG08579.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA5194

  Query Match 59.1%; Score 52; DB 2; Length 267;
  Best Local Similarity 57.1%; Pred. No. 0.23;
  Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SKITHRIHWESASL 15
:|| | :||: |||
Db 118 AKIAHHLHWQHASL 131

RESULT 4
C3RT
complement C3 precursor - rat
N;Alternate names: 37K phospholipase A2 inhibitory protein
N;Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) C3b subuni
C;Species: Rattus norvegicus (Norway rat)
C;Date: 04-Dec-1992 #sequence_revision 07-Oct-1994 #text_change 18-Jun-1999
C;Accession: S15764; A54562; A01260; B35979; A35979; PN0567; PN0566; A32281; S08692
R;Misumi, Y.; Sohda, M.; Ikehara, Y.
Nucleic Acids Res. 18, 2178, 1990
A;Title: Nucleotide and deduced amino acid sequence of rat complement C3.
A;Reference number: S15764; MUID:90245672; PMID:2336397
A;Accession: S15764
A;Molecule type: mRNA
A;Residues: 1-1663 <MIS>
A;Cross-references: EMBL:X52477; NID:g56953; PIDN:CAA36716.1; PID:g56954
R;Sundstrom, S.A.; Komm, B.S.; Ponce-de-Leon, H.; Yi, Z.; Teuscher, C.; Lyttle, C.R.
J. Biol. Chem. 264, 16941-16947, 1989
A;Title: Estrogen regulation of tissue-specific expression of complement C3.
A;Reference number: A54562; MUID:89380332; PMID:2674144
A;Accession: A54562
A;Status: translation not shown
A;Molecule type: mRNA

A;Residues: 'P',1316-1595 <SUN>
A;Cross-references: GB:M29866; NID:g203200; PIDN:AAA40837.1; PID:g554423
R;Jacobs, J.W.; Rubin, J.S.; Hugli, T.E.; Bogardt, R.A.; Mariz, I.K.; Daniels, J.S.; Dau
Biochemistry 17, 5031-5038, 1978
A;Title: Purification, characterization, and amino acid sequence of rat anaphylatoxin (C
A;Reference number: A01260; MUID:79062262; PMID:309768
A;Accession: A01260
A;Molecule type: protein
A;Residues: 671-703,'K',705-720,'KL',723-748 <JAC>
A;Note: three disulfide bonds are present
R;Suwa, Y.; Kudo, I.; Imaizumi, A.; Okada, M.; Kamimura, T.; Suzuki, Y.; Chang, H.W.; Ha
Proc. Natl. Acad. Sci. U.S.A. 87, 2395-2399, 1990
A;Title: Proteinaceous inhibitors of phospholipase A-2 purified from inflammatory sites
A;Reference number: A35979; MUID:90207203; PMID:2320562
A;Accession: B35979
A;Status: preliminary
A;Molecule type: protein
A;Residues: 'X',998-1005 <SUW>
A;Accession: A35979
A;Molecule type: protein
A;Residues: 'X',961-962,'P',964-969 <SU2>
R;Nakagawa, H.; Komorita, N.
Biochem. Biophys. Res. Commun. 194, 1181-1187, 1993
A;Title: Complement component C3-derived neutrophil chemotactic factors purified from ex
A;Reference number: PN0566; MUID:93356786; PMID:8352775
A;Accession: PN0567
A;Molecule type: protein
A;Residues: 568-592 <NAK>
A;Note: amino end of a C3-derived peptide designated exudate neutrophil chemotactic fact
A;Accession: PN0566
A;Molecule type: protein
A;Residues: 671-687 <NA2>
A;Note: amino end of peptide designated neutrophil chemotactic factor 1 and probably ide
R;Kuivanen, P.C.; Capulong, R.B.; Harkins, R.N.; DeSombre, E.R.
Biochem. Biophys. Res. Commun. 158, 898-905, 1989
A;Title: The estrogen-responsive 110K and 74K rat uterine secretory proteins are structu
A;Reference number: A32281; MUID:89149812; PMID:2645873
A;Accession: A32281
A;Molecule type: protein
A;Residues: 25-41 <KUI>
A;Experimental source: 17beta-estradiol-stimulated uterus of immature rat
A;Note: the authors treat this 74K uterine secretory protein, identical as far as sequen
ent
C;Comment: Complement C3 contains two chains, formed by removal of four residues and lin
alternative complement pathways, releases the C3a anaphylatoxin from the amino end of t
rnative-complement-pathway C3/C5 convertase.
C;Comment: C3a anaphylatoxin is a vasoactive peptide and a mediator of inflammation.
C;Comment: C3b, with its highly reactive thiol group, binds to the surface of foreign pa
e classical-complement-pathway C3/C5 convertase. The activity of C3b is regulated by prc
C;Comment: The major site of synthesis of this plasma protein is the liver.
C;Superfamily: alpha-2-macroglobulin
C;Keywords: acute phase; chemotaxis; complement alternate pathway; complement pathway; g
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-666/Product: complement C3 and C3b beta chain #status predicted <C3BB>
F;25-666,671-1663/Product: complement C3 #status predicted <CC3>
F;25-666,749-1663/Product: complement C3b #status predicted <C3B>
F;671-1663/Product: complement C3 alpha chain #status predicted <CC3A>
F;671-748/Product: C3a anaphylatoxin #status experimental <C3T>
F;749-1663/Product: complement C3b alpha' chain #status predicted <C3BA>
F;946-1303/Product: C3dk fragment #status predicted <CDK>
F;1002-1303/Product: C3d fragment #status predicted <C3D>
F;1424-1457/Region: properdin binding
F;558-816,626-661,693-720,694-727,707-728,873-1513,1101-1158,1358-1489,1389-1458,1506-15
F;748-749/Cleavage site: Arg-Ser (C3 convertase) #status predicted
F;939,1617/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1010-1013/Cross-link: thiolester (Cys-Gln) #status predicted
F;1303-1304/Cleavage site: Arg-Ser (complement factor I) #status predicted
F;1320-1321/Cleavage site: Arg-Ser (complement factor I) #status predicted

Query Match 52.3%; Score 46; DB 1; Length 1663;
Best Local Similarity 58.8%; Pred. No. 19;
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 SSKITHRIHWESASLLR 17
|| |: ||| ||||
Db 1304 SSPTVFRLLWESGSLLR 1320

RESULT 5
H83239
pseudouridine synthase RluA PA3246 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83239
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83239
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-211 <STO>
A;Cross-references: GB:AE004747; GB:AE004091; NID:g9949362; PIDN:AAG06634.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: rluA; PA3246

Query Match 51.1%; Score 45; DB 2; Length 211;
Best Local Similarity 40.0%; Pred. No. 2.9;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SKITHRIHWESASLL 16
::| ||: ||:: |:
Db 50 ARIVHRLDWETSGLM 64

RESULT 6
F75508
mrr restriction system protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: F75508
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: F75508
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-336 <WHI>
A;Cross-references: GB:AE001910; GB:AE000513; NID:g6458198; PIDN:AAF10088.1; PID:g645
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0508
A;Map position: 1

Query Match 51.1%; Score 45; DB 2; Length 336;
Best Local Similarity 50.0%; Pred. No. 4.8;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SKITHRIHWESASLLR 17
||: ||| | ::| |
Db 72 SKVRHRIAWACSNLYR 87

RESULT 7
C3MS
complement C3 precursor - mouse
N;Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) C3b subuni
C;Species: Mus musculus (house mouse)
C;Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 18-Jun-1999
C;Accession: A92459; B92459; A92460; A93938; A21898; A54561; S16369; S16189; I49563;

R;Lundwall, A.; Wetsel, R.A.; Domdey, H.; Tack, B.F.; Fey, G.H.
J. Biol. Chem. 259, 13851-13856, 1984
A;Title: Structure of murine complement component C3: I. Nucleotide sequence of cloned c
A;Reference number: A92459; MUID:85054818; PMID:6548745
A;Accession: A92459
A;Molecule type: mRNA
A;Residues: 1-724 <LU1>
A;Accession: B92459
A;Molecule type: DNA
A;Residues: 1-124 <LU2>
R;Wetsel, R.A.; Lundwall, A.; Davidson, F.; Gibson, T.; Tack, B.F.; Fey, G.H.
J. Biol. Chem. 259, 13857-13862, 1984
A;Title: Structure of murine complement component C3: II. Nucleotide sequence of cloned
A;Reference number: A92460; MUID:85054819; PMID:6094532
A;Accession: A92460
A;Molecule type: mRNA
A;Residues: 671-1663 <WET>
R;Domdey, H.; Wiebauer, K.; Kazmaier, M.; Muller, V.; Odink, K.; Fey, G.
Proc. Natl. Acad. Sci. U.S.A. 79, 7619-7623, 1982
A;Title: Characterization of the mRNA and cloned cDNA specifying the third component of
A;Reference number: A93938; MUID:83117730; PMID:6961437
A;Contents: C3a
A;Accession: A93938
A;Molecule type: mRNA
A;Residues: 671-748 <DOM>
R;Sottrup-Jensen, L.; Stepanik, T.M.; Kristensen, T.; Lonblad, P.B.; Jones, C.M.; Wierzb
Proc. Natl. Acad. Sci. U.S.A. 82, 9-13, 1985
A;Title: Common evolutionary origin of alpha2-macroglobulin and complement components C3
A;Reference number: A21898; MUID:85113177; PMID:2578664
A;Accession: A21898
A;Molecule type: mRNA
A;Residues: 25-1663 <SOT>
R;Hamada, J.; Cavanaugh, P.G.; Miki, K.; Nicolson, G.L.
Cancer Res. 53, 4418-4423, 1993
A;Title: A paracrine migration-stimulating factor for metastatic tumor cells secreted by
A;Reference number: A54561; MUID:93373334; PMID:8364938
A;Accession: A54561
A;Molecule type: protein
A;Residues: 25-41;749-760 <HAM>
A;Experimental source: migration-stimulating factor purified from medium conditioned by
R;Sato, T.; Hong, M.H.; Jin, C.H.; Ishimi, Y.; Udagawa, N.; Shinki, T.; Abe, E.; Suda, T
FEBS Lett. 285, 21-24, 1991
A;Title: The specific production of the third component of complement by osteoblastic ce
A;Reference number: S16189; MUID:91293304; PMID:2065778
A;Accession: S16369
A;Molecule type: protein
A;Residues: 25-31 <SAT>
A;Accession: S16189
A;Status: preliminary
A;Molecule type: protein
A;Residues: 671-677,'X',679-680 <SA2>
R;Fey, G.; Domdey, H.; Wiebauer, K.; Whitehead, A.S.; Odink, K.
Springer Semin. Immunopathol. 6, 119-147, 1983
A;Title: Structure and expression of the C3 gene.
A;Reference number: I49563; MUID:84045280; PMID:6356427
A;Accession: I49563
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 25-136,'Q',138-240 <FEY>
A;Cross-references: GB:M35659; NID:g192280; PIDN:AAA37339.1; PID:g192281
R;Fey, G.H.; Wiebauer, K.; Domdey, H.
Ann. N. Y. Acad. Sci. 421, 307-312, 1983
A;Title: Amino acid sequences of mouse complement C3 derived from nucleotide sequences c
A;Reference number: I49576; MUID:84201365; PMID:6609661
A;Accession: I49576
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 658-761 <RES>
A;Cross-references: GB:M33032; NID:g192391; PIDN:AAA37378.1; PID:g192392
C;Comment: Complement C3 contains two chains, formed by removal of four residues and lin
 alternative complement pathways, releases the C3a anaphylatoxin from the amino end of t
rnative-complement-pathway C3/C5 convertase.
C;Comment: C3a anaphylatoxin is a vasoactive peptide and a mediator of inflammation.
C;Comment: C3b, with its highly reactive thiol group, binds to the surface of foreign
e classical-complement-pathway C3/C5 convertase. The activity of C3b is regulated by
C;Comment: The major site of synthesis of this plasma protein is the liver.
C;Genetics:
A;Introns: 27/2; 90/3
A;Note: the list of introns may be incomplete
C;Superfamily: alpha-2-macroglobulin
C;Keywords: acute phase; complement alternate pathway; complement pathway; glycoprote
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-666/Product: complement C3 and C3b beta chain #status predicted <C3BB>
F;25-666,671-1663/Product: complement C3 #status predicted <CC3>
F;25-666,749-1663/Product: C3b #status predicted <C3B>
F;671-1663/Product: complement C3 alpha chain #status predicted <CC3A>
F;671-748/Product: C3a anaphylatoxin #status predicted <C3T>
F;749-1663/Product: C3b alpha' chain #status predicted <C3BA>
F;946-1303/Product: C3dk fragment #status predicted <CDK>
F;1002-1303/Product: C3d fragment #status predicted <C3D>
F;1424-1457/Region: properdin binding
F;559-816,626-661,693-720,694-727,707-728,873-1513,1101-1158,1358-1489,1389-1458,1506
F;748-749/Cleavage site: Arg-Ser (C3 convertase) #status predicted
F;939,1617/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1010-1013/Cross-link: thiolester (Cys-Gln) #status predicted
F;1303-1304/Cleavage site: Arg-Ser (complement factor I) #status predicted
F;1320-1321/Cleavage site: Arg-Ser (complement factor I) #status predicted

  Query Match 51.1%; Score 45; DB 1; Length 1663;
  Best Local Similarity 52.9%; Pred. No. 29;
  Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SSKITHRIHWESASLLR 17
Db 1304 SSATTFRLLWENGNLLR 1320

RESULT 8
S67037
SMP3 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein O3527; protein YOR149c
C;Species: Saccharomyces cerevisiae
C;Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 21-Jul-2000
C;Accession: S67037; S13750
R;Bordonne, R.; Camasses, A.; Madania, A.; Martin, R.P.; Poch, O.; Tarassov, I.A.; Wi
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67032
A;Accession: S67037
A;Molecule type: DNA
A;Residues: 1-516 <BOR>
A;Cross-references: EMBL:Z75057; NID:g1420374; PID:e252038; PID:g1420375; MIPS:YOR149
A;Experimental source: strain S288C
R;Irie, K.; Araki, H.; Oshima, Y.
Mol. Gen. Genet. 225, 257-265, 1991
A;Title: Mutations in a Saccharomyces cerevisiae host showing increased holding stabi
A;Reference number: S13750; MUID:91172125; PMID:2005867
A;Accession: S13750
A;Molecule type: DNA
A;Residues: 1-121,'IK',124-162,'G',164-168,'R',170-278,'L',280-516 <IRI>
A;Cross-references: EMBL:X58121; NID:g4497; PIDN:CAA41123.1; PID:g4498
C;Genetics:
A;Gene: SGD:SMP3
A;Cross-references: SGD:S0005675; MIPS:YOR149c
A;Map position: 15R
C;Keywords: transmembrane protein
F;9-25/Domain: transmembrane #status predicted <TM1>
F;189-205/Domain: transmembrane #status predicted <TM2>
F;215-231/Domain: transmembrane #status predicted <TM3>
F;271-287/Domain: transmembrane #status predicted <TM4>
F;344-360/Domain: transmembrane #status predicted <TM5>

  Query Match 50.0%; Score 44; DB 2; Length 516;
  Best Local Similarity 63.6%; Pred. No. 12;
  Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 6 HRIHWESASLL 16

```
          :|:||:| |||
Db    207 YRVHWKSFSLL 217


RESULT 9
AH0011
ferredoxin-NADP reductase (EC 1.18.1.2) [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 03-Jun-2002
C;Accession: AH0011
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AH0011
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-248 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC88954.1; PID:g15978201; GSPDB:GN00175
C;Genetics:
A;Gene: fpr
C;Keywords: oxidoreductase

  Query Match             47.7%;  Score 42;  DB 2;  Length 248;
  Best Local Similarity   61.5%;  Pred. No. 11;
  Matches    8;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESA 13
          | ||||  ||  |
Db      6 SGKITHIEHWTDA 18


RESULT 10
C86317
protein T10O22.23 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: C86317
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: C86317
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-280 <STO>
A;Cross-references: GB:AE005172; NID:g8671774; PIDN:AAF78380.1; GSPDB:GN00141
C;Genetics:
A;Gene: T10O22.23
A;Map position: 1

  Query Match             47.7%;  Score 42;  DB 2;  Length 280;
  Best Local Similarity   50.0%;  Pred. No. 13;
  Matches    8;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESASLL 16
          ||  |:|: ||:  ||
Db    122 SSDSTNRLSWENCDLL 137


RESULT 11
E82521
hypothetical protein XF2735 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: E82521
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: E82521
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-401 <SIM>
A;Cross-references: GB:AE004080; GB:AE003849; NID:g9107971; PIDN:AAF85520.1; GSPDB:GN
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF2735

  Query Match             47.7%;  Score 42;  DB 2;  Length 401;
  Best Local Similarity   45.5%;  Pred. No. 20;
  Matches    5;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 ITHRIHWESAS 14
          : ||:||:  |
Db    334 LAHRVHWDEES 344


RESULT 12
G75580
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 15-Jun-2001
C;Accession: G75580
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: G75580
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-474 <WHI>
A;Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12485.1; PID:g646
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRA0272
A;Map position: 2
C;Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF0821

  Query Match             47.7%;  Score 42;  DB 2;  Length 474;
  Best Local Similarity   46.2%;  Pred. No. 24;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESA 13
          |:::| |:||  |
Db    396 SARLTSRLHWRPA 408


RESULT 13
T18946
```

probable phospholipase activating protein C05C10.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 07-Dec-1999
C;Accession: T18946; T24252
R;Matthews, P.
submitted to the EMBL Data Library, February 1995
A;Reference number: Z19049
A;Accession: T18946
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-858 <WIL>
A;Cross-references: EMBL:Z48178; PIDN:CAA88206.1; GSPDB:GN00020; CESP:C05C10.6
A;Experimental source: clone C05C10
R;Wilkinson, J.
submitted to the EMBL Data Library, October 1995
A;Reference number: Z19863
A;Accession: T24252
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-858 <WI2>
A;Cross-references: EMBL:Z66515; PIDN:CAA91354.1; GSPDB:GN00020; CESP:C05C10.6
A;Experimental source: clone R53
C;Genetics:
A;Gene: CESP:C05C10.6
A;Map position: 2
A;Introns: 15/3; 120/1; 155/3; 407/3; 513/1; 549/1; 593/3; 711/2; 786/3; 821/3

  Query Match 47.7%; Score 42; DB 2; Length 858;
  Best Local Similarity 64.3%; Pred. No. 46;
  Matches 9; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

Qy 6 HRIHWE--SASLLR 17
 | |||: |||:||
Db 226 HIIHWDVASASILR 239

RESULT 14
JQ0393
nodulation protein nodA - Azorhizobium caulinodans
N;Alternate names: hypothetical 24.9K protein
C;Species: Azorhizobium caulinodans
A;Note: host Sesbania rostrata
C;Date: 07-Sep-1990 #sequence_revision 27-Jan-1995 #text_change 16-Jul-1999
C;Accession: JQ0393
R;Goethals, K.; Gao, M.; Tomekpe, K.; Van Montagu, M.; Holsters, M.
Mol. Gen. Genet. 219, 289-298, 1989
A;Title: Common nodABC genes in Nod locus 1 of Azorhizobium caulinodans: nucleotide sequ
A;Reference number: JQ0393; MUID:90136519; PMID:2615763
A;Accession: JQ0393
A;Molecule type: DNA
A;Residues: 1-226 <GOE>
A;Cross-references: GB:L18897; NID:g1293899; PIDN:AAB51162.1; PID:g310292
A;Experimental source: strain ORS571
C;Comment: This is one of the proteins, coded by nodulation genes, that are required for
C;Genetics:
A;Gene: nodA
C;Superfamily: nodulation protein nodA
C;Keywords: nodulation

  Query Match 46.6%; Score 41; DB 1; Length 226;
  Best Local Similarity 63.6%; Pred. No. 15;
  Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SKITHRIHWES 12
 ||:| |: |||
Db 33 SKVTWRVAWES 43

RESULT 15
AI3289
hypothetical cytosolic protein BMEI0303 [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C;Accession: AI3289
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A;Reference number: AD3252; PMID:11756688
A;Accession: AI3289
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAL51484.1; PID:g17982196; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI0303
A;Map position: I
C;Superfamily: Rickettsia prowazekii hypothetical protein RP073

  Query Match 46.6%; Score 41; DB 2; Length 229;
  Best Local Similarity 53.8%; Pred. No. 16;
  Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 KITHRIHWESASL 15
 :| :| || ||:|
Db 136 QIRNRTHWNSANL 148

Search completed: February 24, 2003, 15:34:45
Job time : 16 secs

OM protein - protein search, using sw model

Run on: February 24, 2003, 15:32:46 ; Search time 29 Seconds
(without alignments)
120.786 Million cell updates/sec

Title: US-09-846-346-1
Perfect score: 88
Sequence: 1 SSKITHRIHWESASLLR 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	61	69.3	154	6	Q29289	Q29289 sus scrofa
2	61	69.3	1661	6	Q9GKP1	Q9gkp1 sus scrofa
3	60	68.2	167	6	Q9N0M4	Q9n0m4 cervus nipp
4	60	68.2	349	6	O46544	O46544 ovis aries
5	52	59.1	267	16	Q9HTZ5	Q9htz5 pseudomonas
6	46	52.3	441	5	Q8T3J9	Q8t3j9 drosophila
7	45	51.1	211	16	Q9HYZ4	Q9hyz4 pseudomonas
8	45	51.1	336	16	Q9RX07	Q9rx07 deinococcus
9	43	48.9	75	6	Q9GMH7	Q9gmh7 macaca fasc
10	42	47.7	248	16	Q8ZJK6	Q8zjk6 yersinia pe
11	42	47.7	280	10	Q9LM24	Q9lm24 arabidopsis
12	42	47.7	358	4	Q96LL0	Q96ll0 homo sapien
13	42	47.7	360	10	Q9LPP7	Q9lpp7 arabidopsis
14	42	47.7	401	16	Q9P9Y5	Q9p9y5 xylella fas
15	42	47.7	407	5	Q8SYY7	Q8syy7 drosophila
16	42	47.7	411	5	Q9V4I4	Q9v4i4 drosophila
17	42	47.7	474	16	Q9RYN8	Q9ryn8 deinococcus
18	42	47.7	858	5	Q17647	Q17647 caenorhabdi
19	42	47.7	860	5	Q95NM4	Q95nm4 caenorhabdi
20	41.5	47.2	381	2	Q8RTQ7	Q8rtq7 thermodesul
21	41.5	47.2	382	2	Q93EV7	Q93ev7 thermodesul
22	41	46.6	197	17	Q9HK18	Q9hk18 thermoplasm
23	41	46.6	219	13	Q90YC5	Q90yc5 brachydanio
24	41	46.6	229	16	Q8YIY6	Q8yiy6 brucella me
25	41	46.6	318	2	Q9X5J4	Q9x5j4 mycobacteri
26	41	46.6	329	16	Q989X9	Q989x9 rhizobium l
27	41	46.6	386	2	Q9AEX8	Q9aex8 treponema h
28	41	46.6	531	10	Q8W071	Q8w071 oryza sativ
29	41	46.6	541	16	Q9A017	Q9a017 streptococc
30	41	46.6	615	16	Q9CHM3	Q9chm3 lactococcus
31	41	46.6	1585	16	Q8UBT4	Q8ubt4 agrobacteri
32	40.5	46.0	1417	16	Q8X6G3	Q8x6g3 escherichia
33	40	45.5	191	12	Q9E348	Q9e348 maize necro
34	40	45.5	205	5	Q9NDY6	Q9ndy6 leishmania
35	40	45.5	232	16	Q92KX1	Q92kx1 rhizobium m
36	40	45.5	272	16	Q984A5	Q984a5 rhizobium l
37	40	45.5	274	11	Q9D912	Q9d912 mus musculu
38	40	45.5	285	5	Q18611	Q18611 caenorhabdi
39	40	45.5	286	16	Q98BF5	Q98bf5 rhizobium l
40	40	45.5	332	2	O82937	O82937 escherichia
41	40	45.5	343	2	Q9ZGU3	Q9zgu3 escherichia
42	40	45.5	420	5	Q9VRZ4	Q9vrz4 drosophila
43	40	45.5	468	5	Q969A8	Q969a8 toxoplasma
44	40	45.5	593	16	Q92GD6	Q92gd6 rickettsia
45	40	45.5	879	5	Q9U475	Q9u475 caenorhabdi

ALIGNMENTS

RESULT 1
Q29289
ID Q29289 PRELIMINARY; PRT; 154 AA.
AC Q29289;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Complement C3 (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RX MEDLINE=96327607; PubMed=8672129;
RA Winteroe A.K., Fredholm M., Davies W.;
RT "Evaluation and characterization of a porcine small intestine cDNA
RT library.";
RL Mamm. Genome 7:509-517(1996).
DR EMBL; F14640; CAA23173.1; -.
DR HSSP; P01024; 1C3D.
DR InterPro; IPR001599; MacroglobInA2.
DR Pfam; PF00207; A2M; 1.
FT NON_TER 1 1
FT NON_TER 154 154
SQ SEQUENCE 154 AA; 17440 MW; 6DC7661C1253ED45 CRC64;

Query Match 69.3%; Score 61; DB 6; Length 154;
Best Local Similarity 70.6%; Pred. No. 0.008;
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 SSKITHRIHWESASLLR 17
           |: : ||| ||||||||
Db      97 SAPVRHRILWESASLLR 113
```

RESULT 2
Q9GKP1

```
ID   Q9GKP1      PRELIMINARY;      PRT;  1661 AA.
AC   Q9GKP1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Complement component C3.
GN   C3.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE=21313047; PubMed=11419349;
RA   Wimmers K., Mekchay S., Ponsuksili S., Hardge T., Yerle M.,
RA   Schellander K.;
RT   "Polymorphic sites in exon 15 and 30 of the porcine C3 gene.";
RL   Anim. Genet. 32:46-47(2001).
DR   EMBL; AF154933; AAG40565.1; -.
DR   HSSP; P01024; 1C3D.
DR   InterPro; IPR002890; A2M_N.
DR   InterPro; IPR000020; Anaphylatoxin.
DR   InterPro; IPR001840; Anaphylatoxn.
DR   InterPro; IPR001599; MacroglobInA2.
DR   InterPro; IPR001134; Netrin_C.
DR   Pfam; PF00207; A2M; 1.
DR   Pfam; PF01835; A2M_N; 1.
DR   Pfam; PF01821; ANATO; 1.
DR   Pfam; PF01759; NTR; 1.
DR   PRINTS; PR00004; ANAPHYLATOXN.
DR   ProDom; PD003264; Anaphylatoxin; 1.
DR   SMART; SM00104; ANATO; 1.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
DR   PROSITE; PS01177; ANAPHYLATOXIN_1; 1.
DR   PROSITE; PS01178; ANAPHYLATOXIN_2; 1.
SQ   SEQUENCE   1661 AA;  186806 MW;  4899D0914BE3310C CRC64;

  Query Match             69.3%;  Score 61;  DB 6;  Length 1661;
  Best Local Similarity   70.6%;  Pred. No. 0.1;
  Matches   12;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 SSKITHRIHWESASLLR 17
           |: : ||| ||||||||
Db    1302 SAPVRHRILWESASLLR 1318


RESULT 3
Q9N0M4
ID   Q9N0M4      PRELIMINARY;      PRT;   167 AA.
AC   Q9N0M4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Complement C3 alpha chain (Fragment).
OS   Cervus nippon (Sika deer).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidea;
OC   Cervidae; Cervinae; Cervus.
OX   NCBI_TaxID=9863;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Jiang Y., Sun L.G., Yu Y.L.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF264631; AAF73464.1; -.
DR   HSSP; P01024; 1C3D.
DR   InterPro; IPR001599; MacroglobInA2.
DR   Pfam; PF00207; A2M; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   167 AA;  18671 MW;  12BFE0798290DFA7 CRC64;

  Query Match             68.2%;  Score 60;  DB 6;  Length 167;
  Best Local Similarity   70.6%;  Pred. No. 0.013;
  Matches   12;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 SSKITHRIHWESASLLR 17
           :| : ||| ||||||||
Db      47 NSLVKHRILWESASLLR 63


RESULT 4
O46544
ID   O46544      PRELIMINARY;      PRT;   349 AA.
AC   O46544;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Complement component C3 (Fragment).
GN   C3.
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WHITE ALPINE; TISSUE=LIVER;
RX   MEDLINE=98309471; PubMed=9647256;
RA   Hein W.R., Dudler L., Marston W.L., Landsverk T., Young A.J.,
RA   Avila D.;
RT   "Ubiquitination and dimerization of complement receptor type 2 on
RT   sheep B cells.";
RL   J. Immunol. 161:458-466(1998).
DR   EMBL; AF038130; AAB92374.2; -.
DR   HSSP; P01024; 1C3D.
DR   InterPro; IPR001599; MacroglobInA2.
DR   Pfam; PF00207; A2M; 1.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
FT   NON_TER       1      1
FT   NON_TER     349    349
SQ   SEQUENCE   349 AA;  39679 MW;  70C2023E42ED5EE3 CRC64;

  Query Match             68.2%;  Score 60;  DB 6;  Length 349;
  Best Local Similarity   70.6%;  Pred. No. 0.029;
  Matches   12;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 SSKITHRIHWESASLLR 17
           :| : ||| ||||||||
Db     328 NSLVKHRILWESASLLR 344


RESULT 5
Q9HTZ5
ID   Q9HTZ5      PRELIMINARY;      PRT;   267 AA.
AC   Q9HTZ5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein PA5194.
GN   PA5194.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
```

```
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004932; AAG08579.1; -.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF01569; PAP2; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   267 AA;  30527 MW;  57CD9D2319B6AD7D CRC64;

  Query Match             59.1%;  Score 52;  DB 16;  Length 267;
  Best Local Similarity   57.1%;  Pred. No. 0.52;
  Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 SKITHRIHWESASL 15
          :|| | :||: |||
Db    118 AKIAHHLHWQHASL 131


RESULT 6
Q8T3J9
ID   Q8T3J9      PRELIMINARY;      PRT;   441 AA.
AC   Q8T3J9;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   AT11889p.
GN   CG7196.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY094997; AAM11325.1; -.
SQ   SEQUENCE   441 AA;  52125 MW;  847067D8FA3A3A16 CRC64;

  Query Match             52.3%;  Score 46;  DB 5;  Length 441;
  Best Local Similarity   50.0%;  Pred. No. 9.8;
  Matches    7;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      3 KITHRIHWESASLL 16
          |: |: ||   |||
Db     20 KVVHKNHWRQVSLL 33


RESULT 7
Q9HYZ4
ID   Q9HYZ4      PRELIMINARY;      PRT;   211 AA.
AC   Q9HYZ4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Pseudouridine synthase RluA.
GN   RLUA OR PA3246.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004747; AAG06634.1; -.
DR   InterPro; IPR000613; PseudoU_synth.
DR   InterPro; IPR002990; PSI_RLU.
DR   Pfam; PF00849; PseudoU_synth_2; 1.
DR   ProDom; PD001819; PseudoU_synth; 1.
DR   PROSITE; PS01129; PSI_RLU; 1.
KW   Complete proteome.
SQ   SEQUENCE   211 AA;  24338 MW;  D333B20FCEA55A94 CRC64;

  Query Match             51.1%;  Score 45;  DB 16;  Length 211;
  Best Local Similarity   40.0%;  Pred. No. 6.6;
  Matches    6;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 SKITHRIHWESASLL 16
          ::| ||: ||:: |:
Db     50 ARIVHRLDWETSGLM 64


RESULT 8
Q9RX07
ID   Q9RX07      PRELIMINARY;      PRT;   336 AA.
AC   Q9RX07;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MRR restriction system protein.
GN   DR0508.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001910; AAF10088.1; -.
DR   TIGR; DR0508; -.
KW   Complete proteome.
SQ   SEQUENCE   336 AA;  37335 MW;  E978C50EC4BBC17B CRC64;

  Query Match             51.1%;  Score 45;  DB 16;  Length 336;
  Best Local Similarity   50.0%;  Pred. No. 11;
  Matches    8;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      2 SKITHRIHWESASLLR 17
          ||: ||| |  ::| |
Db     72 SKVRHRIAWACSNLYR 87


RESULT 9
Q9GMH7
ID   Q9GMH7      PRELIMINARY;      PRT;    75 AA.
AC   Q9GMH7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   Hypothetical 8.5 kDa protein.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN PARIETAL LOBE;
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB047973; BAB12384.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   75 AA;  8548 MW;  16A3D3EA2A3DC6AF CRC64;

  Query Match             48.9%;  Score 43;  DB 6;  Length 75;
  Best Local Similarity   50.0%;  Pred. No. 4.8;
  Matches    8;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESASLL 16
          |:|:||  | :: |||
Db     13 SNKMTHHAHCKATSLL 28


RESULT 10
Q8ZJK6
ID   Q8ZJK6      PRELIMINARY;      PRT;   248 AA.
AC   Q8ZJK6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Ferredoxin--NADP reductase (EC 1.18.1.2).
GN   FPR OR YPO0088.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
DR   EMBL; AJ414141; CAC88954.1; -.
DR   InterPro; IPR001834; Cyt_B5_reductase.
DR   InterPro; IPR001433; Oxred_FAD/NAD(P).
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00970; FAD_binding_6; 1.
DR   Pfam; PF00175; NAD_binding; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KW   Oxidoreductase; Complete proteome.
SQ   SEQUENCE   248 AA;  27936 MW;  5D54FA9EE03FDE0E CRC64;

  Query Match             47.7%;  Score 42;  DB 16;  Length 248;
  Best Local Similarity   61.5%;  Pred. No. 26;
  Matches    8;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESA 13
          | ||||  ||  |
Db      6 SGKITHIEHWTDA 18


RESULT 11
Q9LM24
ID   Q9LM24      PRELIMINARY;      PRT;   280 AA.
AC   Q9LM24;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   T10O22.23.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA   Kim C., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,
RA   Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA   Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA   Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA   Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA   Ecker J.R.;
RT   "Genomic sequence for Arabidopsis thaliana BAC T10O22 from chromosome
RT   I.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC069551; AAF78380.1; -.
SQ   SEQUENCE   280 AA;  32809 MW;  774573C5F956FFF7 CRC64;

  Query Match             47.7%;  Score 42;  DB 10;  Length 280;
  Best Local Similarity   50.0%;  Pred. No. 29;
  Matches    8;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESASLL 16
          ||  |:|: ||:  ||
Db    122 SSDSTNRLSWENCDLL 137


RESULT 12
Q96LL0
ID   Q96LL0      PRELIMINARY;      PRT;   358 AA.
AC   Q96LL0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   CDNA FLJ25410 fis, clone TST03087.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA   Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
RA   Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA   Suzuki Y., Hata H., Nakagawa K., Mizuno S., Morinaga M., Kawamura M.,
RA   Sugiyama T., Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A.,
RA   Kawakami B., Nagai K., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK058139; BAB71681.1; -.
DR   InterPro; IPR000038; GTP_Cell_Div.
DR   Pfam; PF00735; GTP_CDC; 1.
DR   ProDom; PD002565; GTP_Cell_Div; 1.
SQ   SEQUENCE   358 AA;  40780 MW;  474DFEF178EEF1E9 CRC64;

  Query Match             47.7%;  Score 42;  DB 4;  Length 358;
  Best Local Similarity   50.0%;  Pred. No. 38;
  Matches    7;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 ITHRIHWESASLLR 17
          ||| ||:|:  ::|
Db    301 ITHNIHYENYRVIR 314


RESULT 13
```

```
Q9LPP7
ID   Q9LPP7      PRELIMINARY;      PRT;   360 AA.
AC   Q9LPP7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   F15H18.23.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA   Kim C., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,
RA   Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA   Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA   Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA   Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA   Ecker J.R.;
RT   "Genomic sequence for Arabidopsis thaliana BAC F15H18 from chromosome
RT   I.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC013354; AAF25991.1; -.
SQ   SEQUENCE   360 AA;  41605 MW;  468AAAFCD8D2749E CRC64;

  Query Match             47.7%;  Score 42;  DB 10;  Length 360;
  Best Local Similarity   50.0%;  Pred. No. 39;
  Matches    8;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 SSKITHRIHWESASLL 16
           ||  |:|: ||:  ||
Db     122 SSDSTNRLSWENCDLL 137


RESULT 14
Q9P9Y5
ID   Q9P9Y5      PRELIMINARY;      PRT;   401 AA.
AC   Q9P9Y5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein Xf2735.
GN   XF2735.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE004080; AAF85520.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   401 AA;  45544 MW;  050ADA91253A6398 CRC64;

  Query Match             47.7%;  Score 42;  DB 16;  Length 401;
  Best Local Similarity   45.5%;  Pred. No. 43;
  Matches    5;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 ITHRIHWESAS 14
           : ||:||:  |
Db     334 LAHRVHWDEES 344


RESULT 15
Q8SYY7
ID   Q8SYY7      PRELIMINARY;      PRT;   407 AA.
AC   Q8SYY7;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   RE27547p.
GN   CG1859.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY071238; AAL48860.1; -.
SQ   SEQUENCE   407 AA;  44863 MW;  5D2A46A75CB6DD78 CRC64;

  Query Match             47.7%;  Score 42;  DB 5;  Length 407;
  Best Local Similarity   77.8%;  Pred. No. 44;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       6 HRIHWESAS 14
           ||: |||||
Db     150 HRLSWESAS 158


Search completed: February 24, 2003, 15:34:23
Job time : 31 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 24, 2003, 15:32:45 ; Search time 35 Seconds
(without alignments)
64.722 Million cell updates/sec

Title: US-09-846-346-1
Perfect score: 88
Sequence: 1 SSKITHRIHWESASLLR 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	88	100.0	1540	22	ABG25976	Novel human diagno
2	88	100.0	1592	18	AAW34623	Human C3 protein m
3	88	100.0	1635	18	AAW34624	Human C3 protein m
4	88	100.0	1657	18	AAW34629	Human C3 protein m
5	88	100.0	1661	18	AAW34625	Human C3 protein m
6	88	100.0	1663	17	AAR94028	Human C3 precursor
7	88	100.0	1663	17	AAR94029	Human modified C3
8	88	100.0	1663	17	AAR94030	Human modified C3
9	88	100.0	1663	18	AAW34619	Human C3 protein m
10	88	100.0	1663	18	AAW34620	Human C3 protein m
11	88	100.0	1663	18	AAW34621	Human C3 protein m
12	88	100.0	1663	18	AAW34627	Human C3 protein m
13	88	100.0	1663	18	AAW34628	Human C3 protein m
14	88	100.0	1663	18	AAW34630	Human C3 protein m
15	88	100.0	1663	18	AAW40989	Human C3 protein m
16	88	100.0	1663	18	AAW40990	Human C3 protein m
17	88	100.0	1663	18	AAW34606	Wild type human C3
18	88	100.0	1663	18	AAW34607	Human C3 protein m
19	88	100.0	1663	18	AAW34610	Human C3 protein m
20	88	100.0	1663	18	AAW34611	Human C3 protein m
21	88	100.0	1663	18	AAW34612	Human C3 protein m
22	88	100.0	1663	18	AAW34613	Human C3 protein m
23	88	100.0	1663	18	AAW34614	Human C3 protein m
24	88	100.0	1663	18	AAW34615	Human C3 protein m
25	88	100.0	1663	18	AAW34616	Human C3 protein m
26	88	100.0	1663	18	AAW34617	Human C3 protein m
27	88	100.0	1663	18	AAW34618	Human C3 protein m
28	88	100.0	1667	18	AAW34626	Human C3 protein m
29	88	100.0	1667	18	AAW34631	Human C3 protein m
30	84	95.5	1663	18	AAW34608	Human C3 protein m
31	84	95.5	1663	18	AAW34609	Human C3 protein m
32	83	94.3	1663	18	AAW40988	Human C3 protein m
33	44	50.0	66	21	AAB23453	Soybean LLS1 prote
34	44	50.0	146	21	AAB33260	Pinus radiata tran
35	44	50.0	563	21	AAP01934	G. max COI1. Glyc
36	44	50.0	563	21	AAB23463	Soybean LLS1 prote
37	43	48.9	72	23	ABP10890	Human ORFX protein
38	43	48.9	280	22	ABB12430	Human bone marrow
39	42	47.7	390	22	ABB96132	Human testicular a
40	42	47.7	390	22	AAM95445	Human reproductive
41	42	47.7	390	22	AAU21691	Novel human neopla
42	42	47.7	390	22	AAU21814	Novel human neopla
43	42	47.7	411	22	ABB58617	Drosophila melanog
44	41	46.6	74	22	AAU15863	Human novel secret
45	41	46.6	119	22	AAO00035	Human polypeptide

ALIGNMENTS

RESULT 1
ABG25976
ID ABG25976 standard; Protein; 1540 AA.
XX
AC ABG25976;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #25967.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS90163.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations

```
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 56335; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1540 AA;

  Query Match             100.0%;  Score 88;  DB 22;  Length 1540;
  Best Local Similarity   100.0%;  Pred. No. 3e-05;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESASLLR 17
          |||||||||||||||||
Db   1304 SSKITHRIHWESASLLR 1320


RESULT 2
AAW34623
ID   AAW34623 standard; Protein; 1592 AA.
XX
AC   AAW34623;
XX
DT   09-APR-1998  (first entry)
XX
DE   Human C3 protein mutant FT-1.
XX
KW   Human; C3 protein; convertase; complement pathway protein; infection;
KW   down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW   complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW   complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1591
FT                   /note= "R1591T mutation"
FT   Misc-difference 1592
FT                   /note= "E1592N mutation"
FT   Misc-difference 1593
FT                   /note= "A1593Stop mutation"
XX
PN   WO9732981-A1.
XX
PD   12-SEP-1997.
XX
PF   04-MAR-1997;   97WO-GB00603.
XX
PR   19-NOV-1996;   96GB-0024028.
PR   07-MAR-1996;   96GB-0004865.
PR   07-JUN-1996;   96GB-0011896.
PR   08-JUL-1996;   96GB-0014293.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1997-457534/42.
XX
PT   Modified complement pathway protein that forms C3 convertase
PT   resistant to down-regulation - used to exhaust the complement
PT   pathway by super-activation, especially for preventing graft
PT   rejection, etc.
XX
PS   Example 17; Page -; 123pp; English.
XX
CC   This sequence represents a mutated human C3 protein of the invention
CC   (see AAW34606 for wild type protein). This protein is a protein of the
CC   invention, and is a modified native complement pathway protein (A) that
CC   forms a down-regulation resistant C3 convertase. (A), their variants,
CC   fragments and conjugates are used to deplete levels of complement
CC   pathway proteins (by superactivation until one or more components are
CC   exhausted), specifically to prevent rejection of foreign material
CC   (particularly a xenograft) but also to prevent complement-mediated
CC   diseases resulting from (surgical) injury or antibody-antigen interaction
CC   in autoimmune disease, also to localise and/or amplify endogenous
CC   complement protein conversion and deposition at a specific site (e.g. a
CC   virus, infected cell or tumour, to increase sensitivity to
CC   complement-mediated responses; a particular application is eliminating
CC   any cancer cells left after surgical removal of a tumour). Also
CC   contemplated is ex vivo treatment, especially by passing blood through a
CC   matrix containing (A) (this may remove additional anaphylactic peptides
CC   and other inflammatory mediators) or killing of leukaemia cells or
CC   MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC   inhibited by factor I, it can bind repeatedly to factor B (which is then
CC   inactivated), causing inactivation of the alternative pathway by
CC   consumption of factor B.
XX
SQ   Sequence   1592 AA;

  Query Match             100.0%;  Score 88;  DB 18;  Length 1592;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-05;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESASLLR 17
          |||||||||||||||||
Db   1304 SSKITHRIHWESASLLR 1320


RESULT 3
AAW34624
ID   AAW34624 standard; Protein; 1635 AA.
XX
AC   AAW34624;
XX
DT   09-APR-1998  (first entry)
XX
DE   Human C3 protein mutant FT-2.
XX
KW   Human; C3 protein; convertase; complement pathway protein; infection;
KW   down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW   complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW   complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1636
FT                   /note= "wild type E mutated to stop codon"
XX
PN   WO9732981-A1.
XX
PD   12-SEP-1997.
XX
PF   04-MAR-1997;   97WO-GB00603.
```

XX
PR 19-NOV-1996; 96GB-0024028.
PR 07-MAR-1996; 96GB-0004865.
PR 07-JUN-1996; 96GB-0011896.
PR 08-JUL-1996; 96GB-0014293.
XX
PA (IMUT-) IMUTRAN LTD.
XX
PI Farries TC, Harrison RA;
XX
DR WPI; 1997-457534/42.
XX
PT Modified complement pathway protein that forms C3 convertase
PT resistant to down-regulation - used to exhaust the complement
PT pathway by super-activation, especially for preventing graft
PT rejection, etc.
XX
PS Example 17; Page -; 123pp; English.
XX
CC This sequence represents a mutated human C3 protein of the invention
CC (see AAW34606 for wild type protein). This protein is a protein of the
CC invention, and is a modified native complement pathway protein (A) that
CC forms a down-regulation resistant C3 convertase. (A), their variants,
CC fragments and conjugates are used to deplete levels of complement
CC pathway proteins (by superactivation until one or more components are
CC exhausted), specifically to prevent rejection of foreign material
CC (particularly a xenograft) but also to prevent complement-mediated
CC diseases resulting from (surgical) injury or antibody-antigen interaction
CC in autoimmune disease, also to localise and/or amplify endogenous
CC complement protein conversion and deposition at a specific site (e.g. a
CC virus, infected cell or tumour, to increase sensitivity to
CC complement-mediated responses; a particular application is eliminating
CC any cancer cells left after surgical removal of a tumour). Also
CC contemplated is ex vivo treatment, especially by passing blood through a
CC matrix containing (A) (this may remove additional anaphylactic peptides
CC and other inflammatory mediators) or killing of leukaemia cells or
CC MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC inhibited by factor I, it can bind repeatedly to factor B (which is then
CC inactivated), causing inactivation of the alternative pathway by
CC consumption of factor B.
XX
SQ Sequence 1635 AA;

Query Match 100.0%; Score 88; DB 18; Length 1635;
Best Local Similarity 100.0%; Pred. No. 3.2e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSKITHRIHWESASLLR 17
||||||||||||||||||
Db 1304 SSKITHRIHWESASLLR 1320

RESULT 4
AAW34629
ID AAW34629 standard; Protein; 1657 AA.
XX
AC AAW34629;
XX
DT 09-APR-1998 (first entry)
XX
DE Human C3 protein mutant FR-2.
XX
KW Human; C3 protein; convertase; complement pathway protein; infection;
KW down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1638..1645
FT /note= "wild type residues QDEENQKQ mutated to SS"
XX
PN WO9732981-A1.
XX
PD 12-SEP-1997.
XX
PF 04-MAR-1997; 97WO-GB00603.
XX
PR 19-NOV-1996; 96GB-0024028.
PR 07-MAR-1996; 96GB-0004865.
PR 07-JUN-1996; 96GB-0011896.
PR 08-JUL-1996; 96GB-0014293.
XX
PA (IMUT-) IMUTRAN LTD.
XX
PI Farries TC, Harrison RA;
XX
DR WPI; 1997-457534/42.
XX
PT Modified complement pathway protein that forms C3 convertase
PT resistant to down-regulation - used to exhaust the complement
PT pathway by super-activation, especially for preventing graft
PT rejection, etc.
XX
PS Example 17; Page -; 123pp; English.
XX
CC This sequence represents a mutated human C3 protein of the invention
CC (see AAW34606 for wild type protein). This protein is a protein of the
CC invention, and is a modified native complement pathway protein (A) that
CC forms a down-regulation resistant C3 convertase. (A), their variants,
CC fragments and conjugates are used to deplete levels of complement
CC pathway proteins (by superactivation until one or more components are
CC exhausted), specifically to prevent rejection of foreign material
CC (particularly a xenograft) but also to prevent complement-mediated
CC diseases resulting from (surgical) injury or antibody-antigen interaction
CC in autoimmune disease, also to localise and/or amplify endogenous
CC complement protein conversion and deposition at a specific site (e.g. a
CC virus, infected cell or tumour, to increase sensitivity to
CC complement-mediated responses; a particular application is eliminating
CC any cancer cells left after surgical removal of a tumour). Also
CC contemplated is ex vivo treatment, especially by passing blood through a
CC matrix containing (A) (this may remove additional anaphylactic peptides
CC and other inflammatory mediators) or killing of leukaemia cells or
CC MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC inhibited by factor I, it can bind repeatedly to factor B (which is then
CC inactivated), causing inactivation of the alternative pathway by
CC consumption of factor B.
XX
SQ Sequence 1657 AA;

Query Match 100.0%; Score 88; DB 18; Length 1657;
Best Local Similarity 100.0%; Pred. No. 3.2e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSKITHRIHWESASLLR 17
||||||||||||||||||
Db 1304 SSKITHRIHWESASLLR 1320

RESULT 5
AAW34625
ID AAW34625 standard; Protein; 1661 AA.
XX
AC AAW34625;
XX
DT 09-APR-1998 (first entry)
XX
DE Human C3 protein mutant FT-3.
XX
KW Human; C3 protein; convertase; complement pathway protein; infection;
KW down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS Homo sapiens.

```
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1607..1614
FT                   /note= "wild type residues LSSDFWGE mutated to KEALQI"
XX
PN   WO9732981-A1.
XX
PD   12-SEP-1997.
XX
PF   04-MAR-1997;    97WO-GB00603.
XX
PR   19-NOV-1996;    96GB-0024028.
PR   07-MAR-1996;    96GB-0004865.
PR   07-JUN-1996;    96GB-0011896.
PR   08-JUL-1996;    96GB-0014293.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1997-457534/42.
XX
PT   Modified complement pathway protein that forms C3 convertase
PT   resistant to down-regulation - used to exhaust the complement
PT   pathway by super-activation, especially for preventing graft
PT   rejection, etc.
XX
PS   Example 17; Page -; 123pp; English.
XX
CC   This sequence represents a mutated human C3 protein of the invention
CC   (see AAW34606 for wild type protein). This protein is a protein of the
CC   invention, and is a modified native complement pathway protein (A) that
CC   forms a down-regulation resistant C3 convertase. (A), their variants,
CC   fragments and conjugates are used to deplete levels of complement
CC   pathway proteins (by superactivation until one or more components are
CC   exhausted), specifically to prevent rejection of foreign material
CC   (particularly a xenograft) but also to prevent complement-mediated
CC   diseases resulting from (surgical) injury or antibody-antigen interaction
CC   in autoimmune disease, also to localise and/or amplify endogenous
CC   complement protein conversion and deposition at a specific site (e.g. a
CC   virus, infected cell or tumour, to increase sensitivity to
CC   complement-mediated responses; a particular application is eliminating
CC   any cancer cells left after surgical removal of a tumour). Also
CC   contemplated is ex vivo treatment, especially by passing blood through a
CC   matrix containing (A) (this may remove additional anaphylactic peptides
CC   and other inflammatory mediators) or killing of leukaemia cells or
CC   MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC   inhibited by factor I, it can bind repeatedly to factor B (which is then
CC   inactivated), causing inactivation of the alternative pathway by
CC   consumption of factor B.
XX
SQ   Sequence   1661 AA;

  Query Match             100.0%;  Score 88;  DB 18;  Length 1661;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-05;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSKITHRIHWESASLLR 17
           |||||||||||||||||
Db    1304 SSKITHRIHWESASLLR 1320


RESULT 6
AAR94028
ID   AAR94028 standard; Protein; 1663 AA.
XX
AC   AAR94028;
XX
DT   21-MAY-1996  (first entry)
XX
DE   Human C3 precursor.
XX
KW   C3 protein; convertase; Factor I; Factor H; complement.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /label= Sig_peptide
FT   Protein         23..667
FT                   /note= "C3 beta chain"
FT   Peptide         668..671
FT                   /note= "amino acids 668-671 are removed when the
FT                          precursor is cleaved into the alpha and
FT                          beta chains"
FT   Protein         672..1663
FT                   /note= "C3 alpha chain"
XX
PN   WO9607738-A2.
XX
PD   14-MAR-1996.
XX
PF   08-SEP-1995;    95WO-GB02121.
XX
PR   04-MAY-1995;    95GB-0009102.
PR   08-SEP-1994;    94GB-0018147.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1996-171613/17.
DR   N-PSDB; AAT17738.
XX
PT   Mutant complement pathway protein forming stable C3 convertase -
PT   for generalised complement depletion or localised complement
PT   activation
XX
PS   Disclosure; Fig 1; 81pp; English.
XX
CC   Human C3 protein (AAR94028) was produced by expression of a cDNA
CC   sequence (AAT17738) isolated from a human liver cDNA library.
CC   C3 is a complement pathway protein that is suceptible to cleavage
CC   by Factor I and is also susceptible to the inhibitory action
CC   of Factor H.  Mutants of C3 (AAR94029 and AAR94030) have been
CC   produced by site-directed mutagenesis.  These mutants can be
CC   used to super-active the complement system, or to induce
CC   localised super-activation at a specific target to increase
CC   the target's sensitivity to complement-mediated destruction.
XX
SQ   Sequence   1663 AA;

  Query Match             100.0%;  Score 88;  DB 17;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-05;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSKITHRIHWESASLLR 17
           |||||||||||||||||
Db    1304 SSKITHRIHWESASLLR 1320


RESULT 7
AAR94029
ID   AAR94029 standard; Protein; 1663 AA.
XX
AC   AAR94029;
XX
DT   21-MAY-1996  (first entry)
XX
DE   Human modified C3 (R1303X).
XX
KW   C3 protein; convertase; Factor I; Factor H; complement;
KW   tumour; infection; therapy.
XX
OS   Synthetic.
XX
```

```
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /label= Sig_peptide
FT   Protein         23..667
FT                   /note= "C3 beta chain"
FT   Peptide         668..671
FT                   /note= "amino acids 668-671 are removed when the
FT                          precursor is cleaved into the alpha and
FT                          beta chains"
FT   Protein         672..1663
FT                   /note= "C3 alpha chain"
FT   Misc-difference 1303
FT                   /label= Glu, Gly, Gln
XX
PN   WO9607738-A2.
XX
PD   14-MAR-1996.
XX
PF   08-SEP-1995;   95WO-GB02121.
XX
PR   04-MAY-1995;   95GB-0009102.
PR   08-SEP-1994;   94GB-0018147.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1996-171613/17.
XX
PT   Mutant complement pathway protein forming stable C3 convertase  -
PT   for generalised complement depletion or localised complement
PT   activation
XX
PS   Claim 8; Fig 1; 81pp; English.
XX
CC   A modified human C3 protein (AAR94029) differs from the wild-type
CC   (AAR94028) by substitution of Arg-1303 by glutamic acid, glycine
CC   or glutamine.  It is obtained by site-directed mutagenesis of
CC   C3-encoding cDNA (AAT17738).  The modification results in improved
CC   resistance to cleavage by Factor I in comparison to wild-type C3.
CC   This allows the modified C3 to be used therapeutically to
CC   super-active the complement system or the increase a target's
CC   (e.g. tumour, pathogen or virus-infected cell) sensitivity to
CC   complement-mediated destruction.
XX
SQ   Sequence   1663 AA;

  Query Match             100.0%;  Score 88;  DB 17;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-05;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESASLLR 17
          |||||||||||||||||
Db   1304 SSKITHRIHWESASLLR 1320


RESULT 8
AAR94030
ID   AAR94030 standard; Protein; 1663 AA.
XX
AC   AAR94030;
XX
DT   21-MAY-1996  (first entry)
XX
DE   Human modified C3 (D752G, E753S, D754G).
XX
KW   C3 protein; convertase; Factor I; Factor H; complement; tumour;
KW   infection; therapy.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /label= Sig_peptide
FT   Protein         23..667
FT                   /note= "C3 beta chain"
FT   Peptide         668..671
FT                   /note= "amino acids 668-671 are removed when the
FT                          precursor is cleaved into the alpha and
FT                          beta chains"
FT   Protein         672..1663
FT                   /note= "C3 alpha chain"
XX
PN   WO9607738-A2.
XX
PD   14-MAR-1996.
XX
PF   08-SEP-1995;   95WO-GB02121.
XX
PR   04-MAY-1995;   95GB-0009102.
PR   08-SEP-1994;   94GB-0018147.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1996-171613/17.
XX
PT   Mutant complement pathway protein forming stable C3 convertase  -
PT   for generalised complement depletion or localised complement
PT   activation
XX
PS   Claim 11; Fig 1; 81pp; English.
XX
CC   A modified human C3 protein (AAR94030) differs from the wild-type
CC   (AAR94028) by substitution of Asp-Glu-Asp at positions 752-754 by
CC   Gly-Ser-Gly.  It is obtained by site-directed mutagenesis of
CC   C3-encoding cDNA (AAT17738).  The modification reduces the
CC   interaction of C3b/C3i with Factor H in comparison to wild-type
CC   C3.  This allows the modified C3 to be used therapeutically to
CC   super-active the complement system or the increase a target's
CC   (e.g. tumour, pathogen or virus-infected cell) sensitivity to
CC   complement-mediated destruction.
XX
SQ   Sequence   1663 AA;

  Query Match             100.0%;  Score 88;  DB 17;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-05;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESASLLR 17
          |||||||||||||||||
Db   1304 SSKITHRIHWESASLLR 1320


RESULT 9
AAW34619
ID   AAW34619 standard; Protein; 1663 AA.
XX
AC   AAW34619;
XX
DT   09-APR-1998  (first entry)
XX
DE   Human C3 protein mutant DV-9.
XX
KW   Human; C3 protein; convertase; complement pathway protein; infection;
KW   down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW   complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW   complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1216
FT                   /note= "D1216G mutation"
FT   Misc-difference 1217
```

```
FT                   /note= "K1217E mutation"
FT   Misc-difference 1218
FT                   /note= "N1218D mutation"
FT   Misc-difference 1219
FT                   /note= "R1219H mutation"
XX
PN   WO9732981-A1.
XX
PD   12-SEP-1997.
XX
PF   04-MAR-1997;    97WO-GB00603.
XX
PR   19-NOV-1996;    96GB-0024028.
PR   07-MAR-1996;    96GB-0004865.
PR   07-JUN-1996;    96GB-0011896.
PR   08-JUL-1996;    96GB-0014293.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1997-457534/42.
XX
PT   Modified complement pathway protein that forms C3 convertase
PT   resistant to down-regulation - used to exhaust the complement
PT   pathway by super-activation, especially for preventing graft
PT   rejection, etc.
XX
PS   Example 14; Page -; 123pp; English.
XX
CC   This sequence represents a mutated human C3 protein of the invention
CC   (see AAW34606 for wild type protein). This protein is a protein of the
CC   invention, and is a modified native complement pathway protein (A) that
CC   forms a down-regulation resistant C3 convertase. (A), their variants,
CC   fragments and conjugates are used to deplete levels of complement
CC   pathway proteins (by superactivation until one or more components are
CC   exhausted), specifically to prevent rejection of foreign material
CC   (particularly a xenograft) but also to prevent complement-mediated
CC   diseases resulting from (surgical) injury or antibody-antigen interaction
CC   in autoimmune disease, also to localise and/or amplify endogenous
CC   complement protein conversion and deposition at a specific site (e.g. a
CC   virus, infected cell or tumour, to increase sensitivity to
CC   complement-mediated responses; a particular application is eliminating
CC   any cancer cells left after surgical removal of a tumour). Also
CC   contemplated is ex vivo treatment, especially by passing blood through a
CC   matrix containing (A) (this may remove additional anaphylactic peptides
CC   and other inflammatory mediators) or killing of leukaemia cells or
CC   MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC   inhibited by factor I, it can bind repeatedly to factor B (which is then
CC   inactivated), causing inactivation of the alternative pathway by
CC   consumption of factor B.
XX
SQ   Sequence   1663 AA;

  Query Match             100.0%;  Score 88;  DB 18;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-05;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSKITHRIHWESASLLR 17
           |||||||||||||||||
Db    1304 SSKITHRIHWESASLLR 1320


RESULT 10
AAW34620
ID   AAW34620 standard; Protein; 1663 AA.
XX
AC   AAW34620;
XX
DT   09-APR-1998  (first entry)
XX
DE   Human C3 protein mutant CV-4.
XX
KW   Human; C3 protein; convertase; complement pathway protein; infection;
KW   down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW   complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW   complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1260
FT                   /note= "R1260N mutation"
FT   Misc-difference 1264
FT                   /note= "G1264E mutation"
XX
PN   WO9732981-A1.
XX
PD   12-SEP-1997.
XX
PF   04-MAR-1997;    97WO-GB00603.
XX
PR   19-NOV-1996;    96GB-0024028.
PR   07-MAR-1996;    96GB-0004865.
PR   07-JUN-1996;    96GB-0011896.
PR   08-JUL-1996;    96GB-0014293.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1997-457534/42.
XX
PT   Modified complement pathway protein that forms C3 convertase
PT   resistant to down-regulation - used to exhaust the complement
PT   pathway by super-activation, especially for preventing graft
PT   rejection, etc.
XX
PS   Example 14; Page -; 123pp; English.
XX
CC   This sequence represents a mutated human C3 protein of the invention
CC   (see AAW34606 for wild type protein). This protein is a protein of the
CC   invention, and is a modified native complement pathway protein (A) that
CC   forms a down-regulation resistant C3 convertase. (A), their variants,
CC   fragments and conjugates are used to deplete levels of complement
CC   pathway proteins (by superactivation until one or more components are
CC   exhausted), specifically to prevent rejection of foreign material
CC   (particularly a xenograft) but also to prevent complement-mediated
CC   diseases resulting from (surgical) injury or antibody-antigen interaction
CC   in autoimmune disease, also to localise and/or amplify endogenous
CC   complement protein conversion and deposition at a specific site (e.g. a
CC   virus, infected cell or tumour, to increase sensitivity to
CC   complement-mediated responses; a particular application is eliminating
CC   any cancer cells left after surgical removal of a tumour). Also
CC   contemplated is ex vivo treatment, especially by passing blood through a
CC   matrix containing (A) (this may remove additional anaphylactic peptides
CC   and other inflammatory mediators) or killing of leukaemia cells or
CC   MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC   inhibited by factor I, it can bind repeatedly to factor B (which is then
CC   inactivated), causing inactivation of the alternative pathway by
CC   consumption of factor B.
XX
SQ   Sequence   1663 AA;

  Query Match             100.0%;  Score 88;  DB 18;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-05;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSKITHRIHWESASLLR 17
           |||||||||||||||||
Db    1304 SSKITHRIHWESASLLR 1320


RESULT 11
AAW34621
ID   AAW34621 standard; Protein; 1663 AA.
```

XX
AC AAW34621;
XX
DT 09-APR-1998 (first entry)
XX
DE Human C3 protein mutant RY-1.
XX
KW Human; C3 protein; convertase; complement pathway protein; infection;
KW down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1427
FT /note= "R1427Q mutation"
FT Misc-difference 1431
FT /note= "K1431D mutation"
FT Misc-difference 1433
FT /note= "E1433Q mutation"
XX
PN WO9732981-A1.
XX
PD 12-SEP-1997.
XX
PF 04-MAR-1997; 97WO-GB00603.
XX
PR 19-NOV-1996; 96GB-0024028.
PR 07-MAR-1996; 96GB-0004865.
PR 07-JUN-1996; 96GB-0011896.
PR 08-JUL-1996; 96GB-0014293.
XX
PA (IMUT-) IMUTRAN LTD.
XX
PI Farries TC, Harrison RA;
XX
DR WPI; 1997-457534/42.
XX
PT Modified complement pathway protein that forms C3 convertase
PT resistant to down-regulation - used to exhaust the complement
PT pathway by super-activation, especially for preventing graft
PT rejection, etc.
XX
PS Example 14; Page -; 123pp; English.
XX
CC This sequence represents a mutated human C3 protein of the invention
CC (see AAW34606 for wild type protein). This protein is a protein of the
CC invention, and is a modified native complement pathway protein (A) that
CC forms a down-regulation resistant C3 convertase. (A), their variants,
CC fragments and conjugates are used to deplete levels of complement
CC pathway proteins (by superactivation until one or more components are
CC exhausted), specifically to prevent rejection of foreign material
CC (particularly a xenograft) but also to prevent complement-mediated
CC diseases resulting from (surgical) injury or antibody-antigen interaction
CC in autoimmune disease, also to localise and/or amplify endogenous
CC complement protein conversion and deposition at a specific site (e.g. a
CC virus, infected cell or tumour, to increase sensitivity to
CC complement-mediated responses; a particular application is eliminating
CC any cancer cells left after surgical removal of a tumour). Also
CC contemplated is ex vivo treatment, especially by passing blood through a
CC matrix containing (A) (this may remove additional anaphylactic peptides
CC and other inflammatory mediators) or killing of leukaemia cells or
CC MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC inhibited by factor I, it can bind repeatedly to factor B (which is then
CC inactivated), causing inactivation of the alternative pathway by
CC consumption of factor B.
XX
SQ Sequence 1663 AA;

Query Match 100.0%; Score 88; DB 18; Length 1663;
Best Local Similarity 100.0%; Pred. No. 3.3e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSKITHRIHWESASLLR 17
||||||||||||||||
Db 1304 SSKITHRIHWESASLLR 1320

RESULT 12
AAW34627
ID AAW34627 standard; Protein; 1663 AA.
XX
AC AAW34627;
XX
DT 09-APR-1998 (first entry)
XX
DE Human C3 protein mutant FT-5.
XX
KW Human; C3 protein; convertase; complement pathway protein; infection;
KW down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1661
FT /note= "C1661S mutation"
XX
PN WO9732981-A1.
XX
PD 12-SEP-1997.
XX
PF 04-MAR-1997; 97WO-GB00603.
XX
PR 19-NOV-1996; 96GB-0024028.
PR 07-MAR-1996; 96GB-0004865.
PR 07-JUN-1996; 96GB-0011896.
PR 08-JUL-1996; 96GB-0014293.
XX
PA (IMUT-) IMUTRAN LTD.
XX
PI Farries TC, Harrison RA;
XX
DR WPI; 1997-457534/42.
XX
PT Modified complement pathway protein that forms C3 convertase
PT resistant to down-regulation - used to exhaust the complement
PT pathway by super-activation, especially for preventing graft
PT rejection, etc.
XX
PS Example 17; Page -; 123pp; English.
XX
CC This sequence represents a mutated human C3 protein of the invention
CC (see AAW34606 for wild type protein). This protein is a protein of the
CC invention, and is a modified native complement pathway protein (A) that
CC forms a down-regulation resistant C3 convertase. (A), their variants,
CC fragments and conjugates are used to deplete levels of complement
CC pathway proteins (by superactivation until one or more components are
CC exhausted), specifically to prevent rejection of foreign material
CC (particularly a xenograft) but also to prevent complement-mediated
CC diseases resulting from (surgical) injury or antibody-antigen interaction
CC in autoimmune disease, also to localise and/or amplify endogenous
CC complement protein conversion and deposition at a specific site (e.g. a
CC virus, infected cell or tumour, to increase sensitivity to
CC complement-mediated responses; a particular application is eliminating
CC any cancer cells left after surgical removal of a tumour). Also
CC contemplated is ex vivo treatment, especially by passing blood through a
CC matrix containing (A) (this may remove additional anaphylactic peptides
CC and other inflammatory mediators) or killing of leukaemia cells or
CC MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC inhibited by factor I, it can bind repeatedly to factor B (which is then
CC inactivated), causing inactivation of the alternative pathway by
CC consumption of factor B.
XX

SQ Sequence 1663 AA;

Query Match 100.0%; Score 88; DB 18; Length 1663;
Best Local Similarity 100.0%; Pred. No. 3.3e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSKITHRIHWESASLLR 17
||||||||||||||||||
Db 1304 SSKITHRIHWESASLLR 1320

RESULT 13
AAW34628
ID AAW34628 standard; Protein; 1663 AA.
XX
AC AAW34628;
XX
DT 09-APR-1998 (first entry)
XX
DE Human C3 protein mutant FR-2.
XX
KW Human; C3 protein; convertase; complement pathway protein; infection;
KW down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1633
FT /note= "E1633R mutation"
FT Misc-difference 1634
FT /note= "E1634D mutation"
FT Misc-difference 1635
FT /note= "D1635T mutation"
FT Misc-difference 1636
FT /note= "E1636T mutation"
XX
PN WO9732981-A1.
XX
PD 12-SEP-1997.
XX
PF 04-MAR-1997; 97WO-GB00603.
XX
PR 19-NOV-1996; 96GB-0024028.
PR 07-MAR-1996; 96GB-0004865.
PR 07-JUN-1996; 96GB-0011896.
PR 08-JUL-1996; 96GB-0014293.
XX
PA (IMUT-) IMUTRAN LTD.
XX
PI Farries TC, Harrison RA;
XX
DR WPI; 1997-457534/42.
XX
PT Modified complement pathway protein that forms C3 convertase
PT resistant to down-regulation - used to exhaust the complement
PT pathway by super-activation, especially for preventing graft
PT rejection, etc.
XX
PS Example 17; Page -; 123pp; English.
XX
CC This sequence represents a mutated human C3 protein of the invention
CC (see AAW34606 for wild type protein). This protein is a protein of the
CC invention, and is a modified native complement pathway protein (A) that
CC forms a down-regulation resistant C3 convertase. (A), their variants,
CC fragments and conjugates are used to deplete levels of complement
CC pathway proteins (by superactivation until one or more components are
CC exhausted), specifically to prevent rejection of foreign material
CC (particularly a xenograft) but also to prevent complement-mediated
CC diseases resulting from (surgical) injury or antibody-antigen interaction
CC in autoimmune disease, also to localise and/or amplify endogenous
CC complement protein conversion and deposition at a specific site (e.g. a
CC virus, infected cell or tumour, to increase sensitivity to
CC complement-mediated responses; a particular application is eliminating
CC any cancer cells left after surgical removal of a tumour). Also
CC contemplated is ex vivo treatment, especially by passing blood through a
CC matrix containing (A) (this may remove additional anaphylactic peptides
CC and other inflammatory mediators) or killing of leukaemia cells or
CC MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC inhibited by factor I, it can bind repeatedly to factor B (which is then
CC inactivated), causing inactivation of the alternative pathway by
CC consumption of factor B.
XX
SQ Sequence 1663 AA;

Query Match 100.0%; Score 88; DB 18; Length 1663;
Best Local Similarity 100.0%; Pred. No. 3.3e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSKITHRIHWESASLLR 17
||||||||||||||||||
Db 1304 SSKITHRIHWESASLLR 1320

RESULT 14
AAW34630
ID AAW34630 standard; Protein; 1663 AA.
XX
AC AAW34630;
XX
DT 09-APR-1998 (first entry)
XX
DE Human C3 protein mutant FR-3.
XX
KW Human; C3 protein; convertase; complement pathway protein; infection;
KW down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1638..1645
FT /note= "wild type residues QDEENQKQ mutated to RSTRQRAA"
XX
PN WO9732981-A1.
XX
PD 12-SEP-1997.
XX
PF 04-MAR-1997; 97WO-GB00603.
XX
PR 19-NOV-1996; 96GB-0024028.
PR 07-MAR-1996; 96GB-0004865.
PR 07-JUN-1996; 96GB-0011896.
PR 08-JUL-1996; 96GB-0014293.
XX
PA (IMUT-) IMUTRAN LTD.
XX
PI Farries TC, Harrison RA;
XX
DR WPI; 1997-457534/42.
XX
PT Modified complement pathway protein that forms C3 convertase
PT resistant to down-regulation - used to exhaust the complement
PT pathway by super-activation, especially for preventing graft
PT rejection, etc.
XX
PS Example 17; Page -; 123pp; English.
XX
CC This sequence represents a mutated human C3 protein of the invention
CC (see AAW34606 for wild type protein). This protein is a protein of the
CC invention, and is a modified native complement pathway protein (A) that
CC forms a down-regulation resistant C3 convertase. (A), their variants,
CC fragments and conjugates are used to deplete levels of complement
CC pathway proteins (by superactivation until one or more components are

```
CC   exhausted), specifically to prevent rejection of foreign material
CC   (particularly a xenograft) but also to prevent complement-mediated
CC   diseases resulting from (surgical) injury or antibody-antigen interaction
CC   in autoimmune disease, also to localise and/or amplify endogenous
CC   complement protein conversion and deposition at a specific site (e.g. a
CC   virus, infected cell or tumour, to increase sensitivity to
CC   complement-mediated responses; a particular application is eliminating
CC   any cancer cells left after surgical removal of a tumour). Also
CC   contemplated is ex vivo treatment, especially by passing blood through a
CC   matrix containing (A) (this may remove additional anaphylactic peptides
CC   and other inflammatory mediators) or killing of leukaemia cells or
CC   MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC   inhibited by factor I, it can bind repeatedly to factor B (which is then
CC   inactivated), causing inactivation of the alternative pathway by
CC   consumption of factor B.
XX
SQ   Sequence   1663 AA;

  Query Match             100.0%;  Score 88;  DB 18;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-05;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESASLLR 17
          |||||||||||||||||
Db   1304 SSKITHRIHWESASLLR 1320


RESULT 15
AAW40989
ID   AAW40989 standard; Protein; 1663 AA.
XX
AC   AAW40989;
XX
DT   09-APR-1998  (first entry)
XX
DE   Human C3 protein mutant R1303X.
XX
KW   Human; C3 protein; convertase; complement pathway protein; infection;
KW   down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW   complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW   complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1303
FT                   /label= Glu, Gln, Gly
XX
PN   WO9732981-A1.
XX
PD   12-SEP-1997.
XX
PF   04-MAR-1997;   97WO-GB00603.
XX
PR   19-NOV-1996;   96GB-0024028.
PR   07-MAR-1996;   96GB-0004865.
PR   07-JUN-1996;   96GB-0011896.
PR   08-JUL-1996;   96GB-0014293.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1997-457534/42.
XX
PT   Modified complement pathway protein that forms C3 convertase
PT   resistant to down-regulation - used to exhaust the complement
PT   pathway by super-activation, especially for preventing graft
PT   rejection, etc.
XX
PS   Claim 8; Page -; 123pp; English.
XX
CC   This sequence represents a mutated human C3 protein of the invention
CC   (see AAW34606 for wild type protein). This protein is a protein of the
CC   invention, and is a modified native complement pathway protein (A) that
CC   forms a down-regulation resistant C3 convertase. (A), their variants,
CC   fragments and conjugates are used to deplete levels of complement
CC   pathway proteins (by superactivation until one or more components are
CC   exhausted), specifically to prevent rejection of foreign material
CC   (particularly a xenograft) but also to prevent complement-mediated
CC   diseases resulting from (surgical) injury or antibody-antigen interaction
CC   in autoimmune disease, also to localise and/or amplify endogenous
CC   complement protein conversion and deposition at a specific site (e.g. a
CC   virus, infected cell or tumour, to increase sensitivity to
CC   complement-mediated responses; a particular application is eliminating
CC   any cancer cells left after surgical removal of a tumour). Also
CC   contemplated is ex vivo treatment, especially by passing blood through a
CC   matrix containing (A) (this may remove additional anaphylactic peptides
CC   and other inflammatory mediators) or killing of leukaemia cells or
CC   MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC   inhibited by factor I, it can bind repeatedly to factor B (which is then
CC   inactivated), causing inactivation of the alternative pathway by
CC   consumption of factor B.
XX
SQ   Sequence   1663 AA;

  Query Match             100.0%;  Score 88;  DB 18;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-05;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESASLLR 17
          |||||||||||||||||
Db   1304 SSKITHRIHWESASLLR 1320


Search completed: February 24, 2003, 15:33:47
Job time : 36 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 24, 2003, 15:32:45 ; Search time 11 Seconds
(without alignments)
64.100 Million cell updates/sec

Title: US-09-846-346-1
Perfect score: 88
Sequence: 1 SSKITHRIHWESASLLR 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	88	100.0	1663	1	CO3_HUMAN	P01024 homo sapien
2	61	69.3	726	1	CO3_RABIT	P12247 oryctolagus
3	46	52.3	1663	1	CO3_RAT	P01026 rattus norv
4	45	51.1	1663	1	CO3_MOUSE	P01027 mus musculu
5	44	50.0	516	1	SMP3_YEAST	Q04174 saccharomyc
6	41	46.6	226	1	NODA_AZOCA	Q07739 azorhizobiu
7	41	46.6	1666	1	CO3_CAVPO	P12387 cavia porce
8	40	45.5	354	1	ALF2_RHOSH	P29271 rhodobacter
9	40	45.5	396	1	RT09_HUMAN	P82933 homo sapien
10	39	44.3	336	1	PTXD_PSEST	O69054 pseudomonas
11	39	44.3	567	1	CYDC_BACSU	P94366 bacillus su
12	39	44.3	790	1	RECA_MYCTU	P26345 mycobacteri
13	39	44.3	851	1	OBP_HSV11	P10193 herpes simp
14	39	44.3	1015	1	TNP3_ECOLI	P03008 escherichia
15	39	44.3	2012	1	DSCA_HUMAN	O60469 homo sapien
16	38	43.2	242	1	YA85_MYCTU	O53433 mycobacteri
17	38	43.2	280	1	GEM2_HUMAN	O14893 homo sapien
18	38	43.2	280	1	YHM7_YEAST	P38790 saccharomyc
19	38	43.2	320	1	NODI_AZOCA	Q07756 azorhizobiu
20	38	43.2	449	1	MYB1_PHYPA	P80074 physcomitre
21	37.5	42.6	261	1	YF25_MYCTU	Q50582 mycobacteri
22	37.5	42.6	608	1	GLMS_YERPE	Q8z9s8 y glucosami
23	37.5	42.6	609	1	GLMS_PASMU	P57963 p glucosami
24	37	42.0	100	1	RL23_BUCAI	P57589 buchnera ap
25	37	42.0	175	1	RMP2_HUMAN	O60895 homo sapien
26	37	42.0	220	1	PRT1_PICAN	P12806 pichia angu
27	37	42.0	238	1	6PGD_PIG	P14332 sus scrofa
28	37	42.0	269	1	GEM2_RAT	Q9qzp1 rattus norv
29	37	42.0	314	1	MIAA_MYCTU	O33232 mycobacteri
30	37	42.0	331	1	PGTB_HUMAN	P53611 homo sapien
31	37	42.0	331	1	PGTB_RAT	Q08603 rattus norv
32	37	42.0	339	1	PGTB_MOUSE	P53612 mus musculu
33	37	42.0	345	1	SEP3_HUMAN	Q9uh03 homo sapien
34	37	42.0	397	1	CD4_ERYPA	Q08339 erythrocebu
35	37	42.0	458	1	CD4_CERAE	Q08338 cercopithec
36	37	42.0	465	1	SEP3_MOUSE	Q9z1s5 mus musculu
37	37	42.0	508	1	YM05_ARCFU	O28078 archaeoglob
38	37	42.0	580	1	SYQ_RALSO	Q8y199 ralstonia s
39	37	42.0	587	1	T9S3_MOUSE	Q9et30 mus musculu
40	37	42.0	589	1	T9S3_HUMAN	Q9hd45 homo sapien
41	37	42.0	698	1	TNPX_ECOLI	Q00042 escherichia
42	37	42.0	751	1	TREA_YEAST	P32356 saccharomyc
43	37	42.0	4385	1	YP73_CAEEL	Q09222 caenorhabdi
44	36.5	41.5	634	1	GLMS_TREPA	O83833 t glucosami
45	36.5	41.5	847	1	VAV3_MOUSE	Q9r0c8 mus musculu

ALIGNMENTS

RESULT 1
CO3_HUMAN
ID CO3_HUMAN STANDARD; PRT; 1663 AA.
AC P01024;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement C3 precursor [Contains: C3a anaphylatoxin].
GN C3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85140166; PubMed=2579379;
RA de Bruijn M.H.L., Fey G.H.;
RT "Human complement component C3: cDNA coding sequence and derived
RT primary structure.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:708-712(1985).
RN [2]
RP SEQUENCE OF 672-748.
RX MEDLINE=76069169; PubMed=1238393;
RA Hugli T.E.;
RT "Human anaphylatoxin (C3a) from the third component of complement.
RT Primary structure.";
RL J. Biol. Chem. 250:8293-8301(1975).
RN [3]
RP SEQUENCE OF 955-966, AND SUBUNITS.
RC TISSUE=Serum;
RX MEDLINE=95293954; PubMed=7539791;
RA Oxvig C., Haaning J., Kristensen L., Wagner J.M., Rubin I.,
RA Stigbrand T., Gleich G.J., Sottrup-Jensen L.;
RT "Identification of angiotensinogen and complement C3dg as novel
RT proteins binding the proform of eosinophil major basic protein in
RT human pregnancy serum and plasma.";
RL J. Biol. Chem. 270:13645-13651(1995).
RN [4]
RP SEQUENCE OF 988-1036.
RX MEDLINE=82174534; PubMed=6175959;
RA Thomas M.L., Janatova J., Gray W.R., Tack B.F.;
RT "Third component of human complement: localization of the internal
RT thiolester bond.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:1054-1058(1982).
RN [5]
RP SEQUENCE OF 1409-1563.
RX MEDLINE=88154452; PubMed=3279119;
RA Daoudaki M.E., Becherer J.D., Lambris J.D.;
RT "A 34-amino acid peptide of the third component of complement
RT mediates properdin binding.";
RL J. Immunol. 140:1577-1580(1988).
RN [6]
RP STRUCTURE BY NMR OF C3A.
RX MEDLINE=88276894; PubMed=3260670;
RA Nettesheim D.G., Edalji R.P., Mollison K.W., Greer J.,
RA Zuiderweg E.R.P.;

```
RT   "Secondary structure of complement component C3a anaphylatoxin in
RT   solution as determined by NMR spectroscopy: differences between
RT   crystal and solution conformations.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:5036-5040(1988).
RN   [7]
RP   MUTAGENESIS OF THIOESTER BOND REGION.
RX   MEDLINE=92250565; PubMed=1577777;
RA   Isaac L., Isenman D.E.;
RT   "Structural requirements for thioester bond formation in human
RT   complement component C3. Reassessment of the role of thioester bond
RT   integrity on the conformation of C3.";
RL   J. Biol. Chem. 267:10062-10069(1992).
RN   [8]
RP   DISULFIDE BONDS.
RX   MEDLINE=93106233; PubMed=8416818;
RA   Dolmer K., Sottrup-Jensen L.;
RT   "Disulfide bridges in human complement component C3b.";
RL   FEBS Lett. 315:85-90(1993).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 996-1303.
RX   MEDLINE=98259089; PubMed=9596584;
RA   Nagar B., Jones R.G., Diefenbach R.J., Isenman D.E., Rini J.M.;
RT   "X-ray crystal structure of C3d: a C3 fragment and ligand for
RT   complement receptor 2.";
RL   Science 280:1277-1281(1998).
RN   [10]
RP   VARIANT C3F/S.
RX   MEDLINE=89309808; PubMed=2473125;
RA   Poznansky M.C., Clissold P.M., Lachmann P.J.;
RT   "The difference between human C3F and C3S results from a single amino
RT   acid change from an asparagine to an aspartate residue at position
RT   1216 on the alpha-chain of the complement component, C3.";
RL   J. Immunol. 143:1254-1258(1989).
RN   [11]
RP   ERRATUM (RETRACTION OF ABOVE ARTICLE).
RX   MEDLINE=90063087; PubMed=2584723;
RA   Poznansky M.C., Clissold P.M., Lachmann P.J.;
RL   J. Immunol. 143:3860-3862(1989).
RN   [12]
RP   VARIANTS GLY-102 AND PRO-314.
RX   MEDLINE=91011240; PubMed=1976733;
RA   Botto M., Yong Fong K., So A.K., Koch C., Walport M.J.;
RT   "Molecular basis of polymorphisms of human complement component C3.";
RL   J. Exp. Med. 172:1011-1017(1990).
RN   [13]
RP   VARIANT ASN-549.
RX   MEDLINE=95050640; PubMed=7961791;
RA   Singer L., Whitehead W.T., Akama H., Katz Y., Fishelson Z.,
RA   Wetsel R.A.;
RT   "Inherited human complement C3 deficiency. An amino acid substitution
RT   in the beta-chain (ASP549 to ASN) impairs C3 secretion.";
RL   J. Biol. Chem. 269:28494-28499(1994).
RN   [14]
RP   VARIANT GLN-1320.
RA   Watanabe Y., Matsui N., Yan K., Nishimukai H., Tokunaga K.,
RA   Juji T., Kobayashi N., Kohsaka T.;
RT   "A novel C3 allotype C3'F02'has an amino acid substitution that may
RT   inhibit iC3b synthesis and cause C3-hypocomplementemia.";
RL   Mol. Immunol. 30:62-62(1993).
CC   -!- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC       COMPLEMENT SYSTEM. ITS PROCESSING BY C3 CONVERTASE IS THE CENTRAL
CC       REACTION IN BOTH CLASSICAL AND ALTERNATIVE COMPLEMENT PATHWAYS.
CC       AFTER ACTIVATION C3B CAN BIND COVALENTLY, VIA ITS REACTIVE
CC       THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR IMMUNE AGGREGATES.
CC   -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C3,
CC       C3A ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC       INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC       PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC       BASOPHILIC LEUKOCYTES.
CC   -!- SUBUNIT: C3 precursor is first processed by the removal of 4 Arg
CC       residues, forming two chains, beta and alpha, linked by a
CC       disulfide bond. C3 convertase activates C3 by cleaving the alpha
CC       chain, releasing C3A anaphylatoxin and generating C3B (beta chain
CC       + alpha' chain). During pregnancy, C3dg exists as a complex
CC       (probably a 2:2:2 heterohexamer) with AGT and the proform of PRG2.
CC   -!- POLYMORPHISM: THERE ARE TWO ALLELES: C3S (C3 SLOW), THE MOST
CC       COMMON ALLELE IN ALL RACES AND C3F (C3 FAST), RELATIVELY FREQUENT
CC       IN CAUCASOIDS, LESS COMMON IN BLACK AMERICAN, EXTREMELY RARE IN
CC       ORIENTALS.
CC   -!- DISEASE: C3 DEFICIENCY CAUSES A SUSCEPTIBILITY TO PYOGENIC
CC       INFECTION.
CC   -!- MISCELLANEOUS: C3B IS RAPIDLY SPLIT IN TWO POSITIONS BY FACTOR I
CC       AND A COFACTOR TO FORM IC3B (INACTIVATED C3B) AND C3F WHICH IS
CC       RELEASED.
CC   -!- MISCELLANEOUS: IC3B IS THE SLOWLY CLEAVED (POSSIBLY BY FACTOR I)
CC       TO FORM C3C AND C3DG. OTHER PROTEASES PRODUCE OTHER FRAGMENTS SUCH
CC       AS C3D OR C3G.
CC   -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC   -!- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K02765; AAA85332.1; -.
DR   PIR; A01257; C3HU.
DR   PIR; A27603; A27603.
DR   PDB; 1C3D; 18-NOV-98.
DR   SWISS-2DPAGE; P01024; HUMAN.
DR   Siena-2DPAGE; P01024; -.
DR   Genew; HGNC:1318; C3.
DR   MIM; 120700; -.
DR   InterPro; IPR002890; A2M_N.
DR   InterPro; IPR000020; Anaphylatoxin.
DR   InterPro; IPR001840; Anaphylatoxn.
DR   InterPro; IPR001599; MacroglobInA2.
DR   InterPro; IPR001134; Netrin_C.
DR   Pfam; PF00207; A2M; 1.
DR   Pfam; PF01759; NTR; 1.
DR   Pfam; PF01821; ANATO; 1.
DR   Pfam; PF01835; A2M_N; 1.
DR   PRINTS; PR00004; ANAPHYLATOXN.
DR   ProDom; PD003264; Anaphylatoxin; 1.
DR   SMART; SM00104; ANATO; 1.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
DR   PROSITE; PS01177; ANAPHYLATOXIN_1; 1.
DR   PROSITE; PS01178; ANAPHYLATOXIN_2; 1.
KW   Complement pathway; Complement alternate pathway; Plasma;
KW   Inflammatory response; Glycoprotein; Signal; Polymorphism;
KW   Disease mutation; 3D-structure.
FT   SIGNAL       1     22
FT   CHAIN       23   1663       COMPLEMENT C3.
FT   CHAIN       23    667       COMPLEMENT C3, BETA CHAIN.
FT   CHAIN      672   1663       COMPLEMENT C3, ALPHA CHAIN.
FT   PEPTIDE    672    748       C3A ANAPHYLATOXIN.
FT   CHAIN      749   1663       C3B ALPHA' CHAIN.
FT   PEPTIDE    749    954       C3C FRAGMENT.
FT   PEPTIDE    955   1303       C3DG FRAGMENT.
FT   PEPTIDE    955   1001       C3G FRAGMENT.
FT   PEPTIDE   1002   1303       C3D FRAGMENT.
FT   PEPTIDE   1304   1320       C3F FRAGMENT.
FT   SITE       748    749       CLEAVAGE (BY C3 CONVERTASE).
FT   SITE       954    955       CLEAVAGE (BY FACTOR I) (POTENTIAL).
FT   SITE      1303   1304       CLEAVAGE (BY FACTOR I).
FT   SITE      1320   1321       CLEAVAGE (BY FACTOR I).
FT   DOMAIN     693    728       ANAPHYLATOXIN-LIKE.
FT   DOMAIN    1424   1456       PROPERDIN-BINDING.
FT   DISULFID   559    816       INTERCHAIN.
FT   DISULFID   627    662
FT   DISULFID   693    720
FT   DISULFID   694    727
FT   DISULFID   707    728
```

```
  Query Match             100.0%;  Score 88;  DB 1;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 6.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESASLLR 17
          |||||||||||||||||
Db   1304 SSKITHRIHWESASLLR 1320


RESULT 2
CO3_RABIT
ID   CO3_RABIT      STANDARD;      PRT;   726 AA.
AC   P12247;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Complement C3 alpha chain (Fragment).
GN   C3.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87006907; PubMed=3019881;
RA   Kusano M., Choi N.H., Tomita M., Yamamoto K., Migita S., Sekiya T.,
RA   Nishimura S.;
RT   "Nucleotide sequence of cDNA and derived amino acid sequence of
RT   rabbit complement component C3 alpha-chain.";
RL   Immunol. Invest. 15:365-378(1986).
CC   -!- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC       COMPLEMENT SYSTEM. ITS PROCESSING BY C3 CONVERTASE IS THE CENTRAL
CC       REACTION IN BOTH CLASSICAL AND ALTERNATIVE COMPLEMENT PATHWAYS.
CC       AFTER ACTIVATION C3B CAN BIND COVALENTLY, VIA ITS REACTIVE
CC       THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR IMMUNE AGGREGATES.
CC   -!- SUBUNIT: C3 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 ARG
CC       RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC       BOND. C3 CONVERTASE ACTIVATES C3 BY CLEAVING THE ALPHA CHAIN,
CC       RELEASING C3A ANAPHYLATOXIN & GENERATING C3B (BETA CHAIN + ALPHA'
CC       CHAIN).
CC   -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M32434; AAA31190.1; -.
DR   PIR; A27602; A27602.
DR   HSSP; P01024; 1C3D.
DR   InterPro; IPR000020; Anaphylatoxin.
DR   InterPro; IPR001599; MacroglobInA2.
DR   InterPro; IPR001134; Netrin_C.
DR   Pfam; PF00207; A2M; 1.
DR   Pfam; PF01759; NTR; 1.
DR   PROSITE; PS01177; ANAPHYLATOXIN_1; PARTIAL.
DR   PROSITE; PS01178; ANAPHYLATOXIN_2; PARTIAL.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
KW   Complement pathway; Complement alternate pathway; Plasma;
KW   Inflammatory response; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN        <1    726       COMPLEMENT C3 ALPHA CHAIN.
FT   THIOLEST     73     76
FT   CARBOHYD      2      2       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    233    233       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    680    680       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   726 AA;  81844 MW;  F4B4C35D461300E9 CRC64;

  Query Match             69.3%;  Score 61;  DB 1;  Length 726;
  Best Local Similarity   70.6%;  Pred. No. 0.0098;
  Matches   12;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESASLLR 17
          || : ||| |:||||||
Db    367 SSPVKHRIVWDSASLLR 383


RESULT 3
CO3_RAT
ID   CO3_RAT        STANDARD;      PRT;  1663 AA.
AC   P01026;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Complement C3 precursor [Contains: C3A anaphylatoxin].
GN   C3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar; TISSUE=Liver;
RX   MEDLINE=90245672; PubMed=2336397;
RA   Misumi Y., Sohda M., Ikehara Y.;
RT   "Nucleotide and deduced amino acid sequence of rat complement C3.";
RL   Nucleic Acids Res. 18:2178-2178(1990).
RN   [2]
RP   SEQUENCE OF 671-748.
RX   MEDLINE=79062262; PubMed=309768;
RA   Jacobs J.W., Rubin J.S., Hugli T.E., Bogardt R.A., Mariz I.K.,
RA   Daniels J.S., Daughaday W.H., Bradshaw R.A.;
RT   "Purification, characterization, and amino acid sequence of rat
RT   anaphylatoxin (C3a).";
RL   Biochemistry 17:5031-5038(1978).
RN   [3]
RP   SEQUENCE OF 1316-1595 FROM N.A.
RX   MEDLINE=89380332; PubMed=2674144;
RA   Sundstrom S.A., Komm B.S., Ponce-De-Leon H., Yi Z., Teuscher C.,
RA   Lyttle C.R.;
RT   "Estrogen regulation of tissue-specific expression of complement C3.";
RL   J. Biol. Chem. 264:16941-16947(1989).
CC   -!- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC       COMPLEMENT SYSTEM. ITS PROCESSING BY C3 CONVERTASE IS THE CENTRAL
CC       REACTION IN BOTH CLASSICAL AND ALTERNATIVE COMPLEMENT PATHWAYS.
CC       AFTER ACTIVATION C3B CAN BIND COVALENTLY, VIA ITS REACTIVE
CC       THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR IMMUNE AGGREGATES.
CC   -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C3,
CC       C3A ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC       INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC       PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC       BASOPHILIC LEUKOCYTES.
CC   -!- SUBUNIT: C3 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 ARG
CC       RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC       BOND. C3 CONVERTASE ACTIVATES C3 BY CLEAVING THE ALPHA CHAIN,
CC       RELEASING C3A ANAPHYLATOXIN & GENERATING C3B (BETA CHAIN + ALPHA'
CC       CHAIN).
CC   -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC   -!- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52477; CAA36716.1; -.
DR   EMBL; M29866; AAA40837.1; ALT_SEQ.
DR   PIR; A01260; A01260.
DR   PIR; S15764; S15764.
```

```
DR   HSSP; P01024; 1C3D.
DR   InterPro; IPR002890; A2M_N.
DR   InterPro; IPR000020; Anaphylatoxin.
DR   InterPro; IPR001840; Anaphylatoxn.
DR   InterPro; IPR001599; MacroglobInA2.
DR   InterPro; IPR001134; Netrin_C.
DR   Pfam; PF00207; A2M; 1.
DR   Pfam; PF01759; NTR; 1.
DR   Pfam; PF01821; ANATO; 1.
DR   Pfam; PF01835; A2M_N; 1.
DR   PRINTS; PR00004; ANAPHYLATOXN.
DR   ProDom; PD003264; Anaphylatoxin; 1.
DR   SMART; SM00104; ANATO; 1.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
DR   PROSITE; PS01177; ANAPHYLATOXIN_1; 1.
DR   PROSITE; PS01178; ANAPHYLATOXIN_2; 1.
KW   Complement pathway; Complement alternate pathway; Plasma;
KW   Inflammatory response; Glycoprotein; Signal.
FT   SIGNAL        1     24
FT   CHAIN        25   1663       COMPLEMENT C3.
FT   CHAIN        25    666       BETA CHAIN.
FT   CHAIN       671   1663       ALPHA CHAIN.
FT   PEPTIDE     671    748       C3A ANAPHYLATOXIN.
FT   CHAIN       749   1663       C3B (ALPHA' CHAIN).
FT   SITE        748    749       CLEAVAGE (BY C3 CONVERTASE).
FT   DOMAIN      693    728       ANAPHYLATOXIN-LIKE.
FT   DISULFID    558    816       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    626    661       BY SIMILARITY.
FT   DISULFID    693    720       BY SIMILARITY.
FT   DISULFID    694    727       BY SIMILARITY.
FT   DISULFID    707    728       BY SIMILARITY.
FT   DISULFID    873   1513       BY SIMILARITY.
FT   DISULFID   1101   1158       BY SIMILARITY.
FT   DISULFID   1358   1489       BY SIMILARITY.
FT   DISULFID   1389   1458       BY SIMILARITY.
FT   DISULFID   1506   1511       BY SIMILARITY.
FT   DISULFID   1518   1590       BY SIMILARITY.
FT   DISULFID   1537   1661       BY SIMILARITY.
FT   THIOLEST   1010   1013
FT   CARBOHYD    939    939       N-LINKED (GLCNAC. . .) (PROBABLE).
FT   CARBOHYD   1617   1617       N-LINKED (GLCNAC. . .) (PROBABLE).
FT   CONFLICT    721    722       LK -> KL (IN REF. 2).
SQ   SEQUENCE   1663 AA;  186460 MW;  2F87CCB143CDD4BC CRC64;

  Query Match             52.3%;  Score 46;  DB 1;  Length 1663;
  Best Local Similarity   58.8%;  Pred. No. 8.3;
  Matches   10;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 SSKITHRIHWESASLLR 17
           ||    |: ||| ||||
Db    1304 SSPTVFRLLWESGSLLR 1320


RESULT 4
CO3_MOUSE
ID   CO3_MOUSE       STANDARD;       PRT;  1663 AA.
AC   P01027;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Complement C3 precursor (HSE-MSF) [Contains: C3A anaphylatoxin].
GN   C3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (LONG ISOFORM).
RX   MEDLINE=85038854; PubMed=6208565;
RA   Fey G.H., Lundwall A., Wetsel R.A., Tack B.F., de Bruijn M.H.L.,
RA   Domdey H.;
RT   "Nucleotide sequence of complementary DNA and derived amino acid
RT   sequence of murine complement protein C3.";
RL   Philos. Trans. R. Soc. Lond., B, Biol. Sci. 306:333-344(1984).
RN   [2]
RP   SEQUENCE OF 671-1663 FROM N.A. (LONG ISOFORM).
RX   MEDLINE=85054819; PubMed=6094532;
RA   Wetsel R.A., Lundwall A., Davidson F., Gibson T., Tack B.F., Fey G.H.;
RT   "Structure of murine complement component C3. II. Nucleotide sequence
RT   of cloned complementary DNA coding for the alpha chain.";
RL   J. Biol. Chem. 259:13857-13862(1984).
RN   [3]
RP   SEQUENCE OF 671-748 FROM N.A.
RX   MEDLINE=83117730; PubMed=6961437;
RA   Domdey H., Wiebauer K., Kazmaier M., Mueller V., Odink K., Fey G.H.;
RT   "Characterization of the mRNA and cloned cDNA specifying the third
RT   component of mouse complement.";
RL   Proc. Natl. Acad. Sci. U.S.A. 79:7619-7623(1982).
RN   [4]
RP   SEQUENCE OF 658-761 FROM N.A.
RX   MEDLINE=84201365; PubMed=6609661;
RA   Fey G.H., Wiebauer K., Domdey H.;
RT   "Amino acid sequences of mouse complement C3 derived from nucleotide
RT   sequences of cloned cDNA.";
RL   Ann. N.Y. Acad. Sci. 421:307-312(1983).
RN   [5]
RP   SEQUENCE OF 1-34 FROM N.A.
RX   MEDLINE=83117622; PubMed=6985486;
RA   Wiebauer K., Domdey H., Diggelmann H., Fey G.;
RT   "Isolation and analysis of genomic DNA clones encoding the third
RT   component of mouse complement.";
RL   Proc. Natl. Acad. Sci. U.S.A. 79:7077-7081(1982).
RN   [6]
RP   SEQUENCE OF 25-41 AND 749-760.
RX   MEDLINE=93373334; PubMed=8364938;
RA   Hamada J.-I., Cavanaugh P.G., Miki K., Nicolson G.L.;
RT   "A paracrine migration-stimulating factor for metastatic tumor cells
RT   secreted by mouse hepatic sinusoidal endothelial cells:
RT   identification as complement component C3b.";
RL   Cancer Res. 53:4418-4423(1993).
RN   [7]
RP   ALTERNATIVE INITIATION.
RX   MEDLINE=95053742; PubMed=7964485;
RA   Cahen-Kramer Y., Martensson I.L., Melchers F.;
RT   "The structure of an alternate form of complement C3 that displays
RT   costimulatory growth factor activity for B lymphocytes.";
RL   J. Exp. Med. 180:2079-2088(1994).
CC   -!- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC       COMPLEMENT SYSTEM. ITS PROCESSING BY C3 CONVERTASE IS THE CENTRAL
CC       REACTION IN BOTH CLASSICAL AND ALTERNATIVE COMPLEMENT PATHWAYS.
CC       AFTER ACTIVATION C3B CAN BIND COVALENTLY, VIA ITS REACTIVE
CC       THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR IMMUNE AGGREGATES.
CC   -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C3,
CC       C3A ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC       INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC       PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC       BASOPHILIC LEUKOCYTES. THE SHORT ISOFORM HAS B-CELL STIMULATORY
CC       ACTIVITY.
CC   -!- SUBUNIT: C3 precursor is first processed by the removal of 4 Arg
CC       residues, forming two chains, beta and alpha, linked by a
CC       disulfide bond. C3 convertase activates C3 by cleaving the alpha
CC       chain, releasing C3A anaphylatoxin and generating C3B (beta chain
CC       + alpha' chain).
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a
CC       short form; are produced by alternative initiation.
CC   -!- MISCELLANEOUS: C3B IS RAPIDLY SPLIT IN TWO POSITIONS BY FACTOR I
CC       AND A COFACTOR TO FORM IC3B (INACTIVATED C3B) AND C3F WHICH IS
CC       RELEASED.
CC   -!- MISCELLANEOUS: IC3B IS THE SLOWLY CLEAVED (POSSIBLY BY FACTOR I)
CC       TO FORM C3C AND C3DG. OTHER PROTEASES PRODUCE OTHER FRAGMENTS SUCH
CC       AS C3D OR C3G.
CC   -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC   -!- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; K02782; AAC42013.1; -.
DR   EMBL; J00369; AAA37336.1; -.
DR   EMBL; J00367; AAA37336.1; JOINED.
DR   EMBL; M33032; AAA37378.1; -.
DR   EMBL; Z37998; CAA86099.2; -.
DR   PIR; A05290; C3MS.
DR   HSSP; P01024; 1C3D.
DR   MGD; MGI:88227; C3.
DR   InterPro; IPR002890; A2M_N.
DR   InterPro; IPR000020; Anaphylatoxin.
DR   InterPro; IPR001840; Anaphylatoxn.
DR   InterPro; IPR001599; MacrogloblnA2.
DR   InterPro; IPR001134; Netrin_C.
DR   Pfam; PF00207; A2M; 1.
DR   Pfam; PF01759; NTR; 1.
DR   Pfam; PF01821; ANATO; 1.
DR   Pfam; PF01835; A2M_N; 1.
DR   PRINTS; PR00004; ANAPHYLATOXN.
DR   ProDom; PD003264; Anaphylatoxin; 1.
DR   SMART; SM00104; ANATO; 1.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
DR   PROSITE; PS01177; ANAPHYLATOXIN_1; 1.
DR   PROSITE; PS01178; ANAPHYLATOXIN_2; 1.
KW   Complement pathway; Complement alternate pathway; Plasma;
KW   Inflammatory response; Glycoprotein; Signal; Alternative initiation.
FT   SIGNAL       1     24
FT   CHAIN       25   1663       COMPLEMENT C3.
FT   CHAIN       25    666       COMPLEMENT C3, BETA CHAIN.
FT   CHAIN      671   1663       COMPLEMENT C3, ALPHA CHAIN.
FT   CHAIN     1129   1663       COMPLEMENT C3, SHORT ISOFORM.
FT   INIT_MET  1129   1129       FOR SHORT ISOFORM.
FT   PEPTIDE    671    748       C3A ANAPHYLATOXIN.
FT   CHAIN      749   1663       C3B (ALPHA' CHAIN).
FT   PEPTIDE    749    954       C3C FRAGMENT.
FT   PEPTIDE    955   1303       C3DG FRAGMENT.
FT   PEPTIDE    955   1001       C3G FRAGMENT.
FT   PEPTIDE   1002   1303       C3D FRAGMENT.
FT   PEPTIDE   1304   1320       C3F FRAGMENT.
FT   SITE       748    749       CLEAVAGE (BY C3 CONVERTASE).
FT   SITE      1303   1304       CLEAVAGE (BY FACTOR I).
FT   SITE      1320   1321       CLEAVAGE (BY FACTOR I).
FT   DOMAIN     693    728       ANAPHYLATOXIN-LIKE.
FT   DISULFID   559    816       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   626    661       BY SIMILARITY.
FT   DISULFID   693    720       BY SIMILARITY.
FT   DISULFID   694    727       BY SIMILARITY.
FT   DISULFID   707    728       BY SIMILARITY.
FT   DISULFID   873   1513       BY SIMILARITY.
FT   DISULFID  1101   1158       BY SIMILARITY.
FT   DISULFID  1358   1489       BY SIMILARITY.
FT   DISULFID  1389   1458       BY SIMILARITY.
FT   DISULFID  1506   1511       BY SIMILARITY.
FT   DISULFID  1518   1590       BY SIMILARITY.
FT   DISULFID  1537   1661       BY SIMILARITY.
FT   DISULFID  1637   1646       BY SIMILARITY.
FT   CARBOHYD   939    939       N-LINKED (GLCNAC. . .).
FT   CARBOHYD  1617   1617       N-LINKED (GLCNAC. . .).
FT   THIOLEST  1010   1013       BY SIMILARITY.
SQ   SEQUENCE  1663 AA;  186482 MW;  DE5546CC769BEA19 CRC64;

  Query Match           51.1%;  Score 45;  DB 1;  Length 1663;
  Best Local Similarity  52.9%;  Pred. No. 12;
  Matches    9;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 SSKITHRIHWESASLLR 17
           ||  | |: ||: :|||
Db    1304 SSATTFRLLWENGNLLR 1320
```

```
RESULT 5
SMP3_YEAST
ID   SMP3_YEAST     STANDARD;      PRT;   516 AA.
AC   Q04174; Q99400;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   SMP3 protein.
GN   SMP3 OR YOR149C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NBW5;
RX   MEDLINE=91172125; PubMed=2005867;
RA   Irie K., Araki H., Oshima Y.;
RT   "Mutations in a Saccharomyces cerevisiae host showing increased
RT   holding stability of the heterologous plasmid pSR1.";
RL   Mol. Gen. Genet. 225:257-265(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY1678;
RA   Ayadi A., Bordonne R., Camasses A., Madania A., Poch O.,
RA   Tarassov I.A., Winsor B., Martin R.P.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ESSENTIAL PROTEIN INVOLVED IN PLASMID MAINTENANCE WITH
CC       SMP2.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: TO S.POMBE SPAC4G8.12C.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X58121; CAA41123.1; -.
DR   EMBL; U55020; AAC49635.1; -.
DR   EMBL; Z75057; CAA99355.1; -.
DR   PIR; S13750; S13750.
DR   SGD; S0005675; SMP3.
KW   Transmembrane.
FT   TRANSMEM      6     26       POTENTIAL.
FT   TRANSMEM     61     81       POTENTIAL.
FT   TRANSMEM    176    196       POTENTIAL.
FT   TRANSMEM    211    231       POTENTIAL.
FT   TRANSMEM    271    291       POTENTIAL.
FT   TRANSMEM    296    316       POTENTIAL.
FT   TRANSMEM    318    338       POTENTIAL.
FT   TRANSMEM    349    369       POTENTIAL.
FT   CONFLICT    122    123       MQ -> IK (IN REF. 1).
FT   CONFLICT    163    163       E -> G (IN REF. 1).
FT   CONFLICT    169    169       S -> R (IN REF. 1).
FT   CONFLICT    279    279       V -> L (IN REF. 1).
SQ   SEQUENCE   516 AA;  59900 MW;  8D8404622CB69534 CRC64;

  Query Match           50.0%;  Score 44;  DB 1;  Length 516;
  Best Local Similarity  63.6%;  Pred. No. 5.2;
  Matches    7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       6 HRIHWESASLL 16
           :|:||:| |||
Db     207 YRVHWKSFSLL 217


RESULT 6
NODA_AZOCA
```

```
ID   NODA_AZOCA     STANDARD;      PRT;   226 AA.
AC   Q07739;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nodulation protein A (EC 2.3.1.-).
GN   NODA.
OS   Azorhizobium caulinodans.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Hyphomicrobium group; Azorhizobium.
OX   NCBI_TaxID=7;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ORS571;
RX   MEDLINE=90136519; PubMed=2615763;
RA   Goethals K., Gao M., Tomekpe K., van Montagu M., Holsters M.;
RT   "Common nodABC genes in Nod locus 1 of Azorhizobium caulinodans:
RT   nucleotide sequence and plant-inducible expression.";
RL   Mol. Gen. Genet. 219:289-298(1989).
CC   -!- FUNCTION: N-ACYLTRANSFERASE REQUIRED FOR NODULATION. ACTS IN THE
CC       PRODUCTION OF A SMALL, HEAT-STABLE COMPOUND (NOD) THAT STIMULATES
CC       MITOSIS IN VARIOUS PLANT PROTOPLASTS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE NODA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L18897; AAB51162.1; -.
DR   PIR; JQ0393; JQ0393.
DR   InterPro; IPR003484; NodA.
DR   Pfam; PF02474; NodA; 1.
DR   PROSITE; PS01349; NODA; 1.
KW   Transferase; Acyltransferase; Nodulation.
SQ   SEQUENCE   226 AA;  24915 MW;  F1992B421A002315 CRC64;

  Query Match             46.6%;  Score 41;  DB 1;  Length 226;
  Best Local Similarity   63.6%;  Pred. No. 7;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 SKITHRIHWES 12
           ||:| |: |||
Db      33 SKVTWRVAWES 43


RESULT 7
CO3_CAVPO
ID   CO3_CAVPO      STANDARD;      PRT;  1666 AA.
AC   P12387;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Complement C3 precursor [Contains: C3A anaphylatoxin].
GN   C3.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90307998; PubMed=1973176;
RA   Auerbach H.S., Burger R., Dodds A., Colten H.R.;
RT   "Molecular basis of complement C3 deficiency in guinea pigs.";
RL   J. Clin. Invest. 86:96-106(1990).
RN   [2]
RP   SEQUENCE OF 676-753.
RX   MEDLINE=89113342; PubMed=3064079;
RA   Gerard N.P., Lively M.O., Gerard C.;
RT   "Amino acid sequence of guinea pig C3a anaphylatoxin.";
RL   Protein Seq. Data Anal. 1:473-478(1988).
RN   [3]
RP   SEQUENCE OF 993-1032.
RX   MEDLINE=83178889; PubMed=6838833;
RA   Thomas M.L., Tack B.F.;
RT   "Identification and alignment of a thiol ester site in the third
RT   component of guinea pig complement.";
RL   Biochemistry 22:942-947(1983).
CC   -!- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC       COMPLEMENT SYSTEM. ITS PROCESSING BY C3 CONVERTASE IS THE CENTRAL
CC       REACTION IN BOTH CLASSICAL AND ALTERNATIVE COMPLEMENT PATHWAYS.
CC       AFTER ACTIVATION C3B CAN BIND COVALENTLY, VIA ITS REACTIVE
CC       THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR IMMUNE AGGREGATES.
CC   -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C3,
CC       C3A ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC       INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC       PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC       BASOPHILIC LEUKOCYTES.
CC   -!- SUBUNIT: C3 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 ARG
CC       RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC       BOND. C3 CONVERTASE ACTIVATES C3 BY CLEAVING THE ALPHA CHAIN,
CC       RELEASING C3A ANAPHYLATOXIN & GENERATING C3B (BETA CHAIN + ALPHA'
CC       CHAIN).
CC   -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC   -!- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M34054; AAA37038.1; -.
DR   PIR; A37156; A37156.
DR   PIR; S03375; S03375.
DR   PIR; D20342; D20342.
DR   HSSP; P01024; 1C3D.
DR   InterPro; IPR002890; A2M_N.
DR   InterPro; IPR000020; Anaphylatoxin.
DR   InterPro; IPR001840; Anaphylatoxn.
DR   InterPro; IPR001599; MacrogloblnA2.
DR   InterPro; IPR001134; Netrin_C.
DR   Pfam; PF00207; A2M; 1.
DR   Pfam; PF01759; NTR; 1.
DR   Pfam; PF01821; ANATO; 1.
DR   Pfam; PF01835; A2M_N; 1.
DR   PRINTS; PR00004; ANAPHYLATOXN.
DR   ProDom; PD003264; Anaphylatoxin; 1.
DR   SMART; SM00104; ANATO; 1.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
DR   PROSITE; PS01177; ANAPHYLATOXIN_1; 1.
DR   PROSITE; PS01178; ANAPHYLATOXIN_2; 1.
KW   Complement pathway; Complement alternate pathway; Plasma;
KW   Inflammatory response; Glycoprotein; Signal.
FT   SIGNAL       1     29
FT   CHAIN       30   1666       COMPLEMENT C3.
FT   CHAIN       30    671       BETA CHAIN.
FT   CHAIN      676   1666       ALPHA CHAIN.
FT   PEPTIDE     676    753       C3A ANAPHYLATOXIN.
FT   CHAIN       754   1666       C3B (ALPHA' CHAIN).
FT   SITE        753    754       CLEAVAGE (BY C3 CONVERTASE).
FT   DISULFID    557    821       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    630    666       BY SIMILARITY.
FT   DOMAIN      698    733       ANAPHYLATOXIN-LIKE.
FT   DISULFID    698    725       BY SIMILARITY.
FT   DISULFID    699    732       BY SIMILARITY.
FT   DISULFID    712    733       BY SIMILARITY.
FT   DISULFID    878   1517       BY SIMILARITY.
FT   DISULFID   1106   1163       BY SIMILARITY.
FT   DISULFID   1363   1493       BY SIMILARITY.
```

```
FT   DISULFID   1394   1462       BY SIMILARITY.
FT   DISULFID   1510   1515       BY SIMILARITY.
FT   DISULFID   1522   1593       BY SIMILARITY.
FT   DISULFID   1540   1664       BY SIMILARITY.
FT   DISULFID   1640   1649       BY SIMILARITY.
FT   THIOLEST   1015   1018
FT   CARBOHYD    944    944       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1620   1620       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    731    731       D -> N (IN REF. 2).
FT   CONFLICT   1013   1013       MISSING (IN REF. 3).
FT   CONFLICT   1018   1018       Q -> E (IN REF. 2).
FT   CONFLICT   1031   1031       MISSING (IN REF. 3).
SQ   SEQUENCE   1666 AA;  186487 MW;  1C1F1219944AFD49 CRC64;

  Query Match             46.6%;  Score 41;  DB 1;  Length 1666;
  Best Local Similarity   52.9%;  Pred. No. 59;
  Matches    9;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESASLLR 17
          ||    |: ||: ||||
Db   1309 SSPSKFRLVWEAGSLLR 1325


RESULT 8
ALF2_RHOSH
ID   ALF2_RHOSH     STANDARD;      PRT;   354 AA.
AC   P29271;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Fructose-bisphosphate aldolase II (EC 4.1.2.13).
GN   CFXB.
OS   Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
OX   NCBI_TaxID=1063;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92041881; PubMed=1939098;
RA   Chen J.-H., Gibson J.L., McCue L.A., Tabita F.R.;
RT   "Identification, expression, and deduced primary structure of
RT   transketolase and other enzymes encoded within the form II CO2
RT   fixation operon of Rhodobacter sphaeroides.";
RL   J. Biol. Chem. 266:20447-20452(1991).
CC   -!- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate = glycerone
CC       phosphate + D-glyceraldehyde 3-phosphate.
CC   -!- COFACTOR: ZINC.
CC   -!- PATHWAY: Glycolysis; sixth step.
CC   -!- PATHWAY: PART OF REDUCTIVE PENTOSE PHOSPHATE PATHWAY OR CALVIN
CC       CYCLE OF PHOTOSYNTHETIC CARBON DIOXIDE ASSIMILATION.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- MISCELLANEOUS: THIS PROTEIN IS ENCODED WITHIN THE FORM II
CC       RIBULOSE-BISPHOSPHATE CARBOXYLASE OPERON.
CC   -!- SIMILARITY: BELONGS TO CLASS II FRUCTOSE-BISPHOSPHATE ALDOLASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M68914; AAA26157.1; -.
DR   PIR; D41080; D41080.
DR   InterPro; IPR000771; F_bP_aldolase.
DR   Pfam; PF01116; F_bP_aldolase; 1.
DR   ProDom; PD002376; F_bP_aldolase; 1.
DR   TIGRFAMs; TIGR00167; cbbA; 1.
DR   PROSITE; PS00602; ALDOLASE_CLASS_II_1; 1.
DR   PROSITE; PS00806; ALDOLASE_CLASS_II_2; 1.
KW   Lyase; Glycolysis; Zinc; Calvin cycle; Multigene family.
FT   METAL        81     81       ZINC (BY SIMILARITY).
FT   METAL        84     84       ZINC (BY SIMILARITY).
SQ   SEQUENCE   354 AA;  38269 MW;  9F547E84FC72ACF5 CRC64;

  Query Match             45.5%;  Score 40;  DB 1;  Length 354;
  Best Local Similarity   33.3%;  Pred. No. 17;
  Matches    5;  Conservative    6;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHWESASL 15
          :::::|  ||  ||:
Db    125 TARVSHMAHWVGASV 139


RESULT 9
RT09_HUMAN
ID   RT09_HUMAN     STANDARD;      PRT;   396 AA.
AC   P82933;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   28S ribosomal protein S9, mitochondrial precursor (MRP-S9).
GN   MRPS9 OR RPMS9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   IDENTIFICATION.
RX   MEDLINE=21276436; PubMed=11279123;
RA   Koc E.C., Burkhart W., Blackburn K., Moseley A., Spremulli L.L.;
RT   "The small subunit of the mammalian mitochondrial ribosome:
RT   identification of the full complement of ribosomal proteins present.";
RL   J. Biol. Chem. 276:19363-19374(2001).
CC   -!- SUBCELLULAR LOCATION: Mitochondrial.
CC   -!- SIMILARITY: BELONGS TO THE S9P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BF034318; -; NOT_ANNOTATED_CDS.
DR   InterPro; IPR000754; Ribosomal_S9.
DR   Pfam; PF00380; Ribosomal_S9; 1.
DR   ProDom; PD001627; Ribosomal_S9; 1.
DR   PROSITE; PS00360; RIBOSOMAL_S9; 1.
KW   Ribosomal protein; Mitochondrion; Transit peptide.
FT   TRANSIT       1      ?       MITOCHONDRION (POTENTIAL).
FT   CHAIN         ?    396       28S RIBOSOMAL PROTEIN S9.
SQ   SEQUENCE   396 AA;  45822 MW;  A4ECC6FD3F7FE9AE CRC64;

  Query Match             45.5%;  Score 40;  DB 1;  Length 396;
  Best Local Similarity   54.5%;  Pred. No. 19;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      6 HRIHWESASLL 16
          |: ||:: |||
Db    175 HQSHWQAKSLL 185


RESULT 10
PTXD_PSEST
ID   PTXD_PSEST     STANDARD;      PRT;   336 AA.
AC   O69054;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
```

```
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phosphonate dehydrogenase (EC 1.20.1.1) (NAD-dependent phosphite
DE   dehydrogenase).
GN   PTXD.
OS   Pseudomonas stutzeri (Pseudomonas perfectomarina).
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=316;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WM88;
RX   MEDLINE=99008986; PubMed=9791102;
RA   Metcalf W.W., Wolfe R.S.;
RT   "Molecular genetic analysis of phosphite and hypophosphite oxidation
RT   by Pseudomonas stutzeri WM88.";
RL   J. Bacteriol. 180:5547-5558(1998).
RN   [2]
RP   SEQUENCE OF 1-15, FUNCTION, ACTIVITY, COFACTOR, ENZYME REGULATION,
RP   SUBUNIT, INDUCTION, AND MASS SPECTROMETRY.
RC   STRAIN=WM88;
RX   MEDLINE=21264507; PubMed=11278981;
RA   Costas A.M.G., White A.K., Metcalf W.W.;
RT   "Purification and characterization of a novel phosphorus-oxidizing
RT   enzyme from Pseudomonas stutzeri WM88.";
RL   J. Biol. Chem. 276:17429-17436(2001).
CC   -!- FUNCTION: Catalyzes phosphite (phosphonate) oxidation.
CC   -!- CATALYTIC ACTIVITY: Phosphonate + NAD(+) + H(2)O = phosphate +
CC       NADH.
CC   -!- ENZYME REGULATION: Inhibited by NaCl, NADH and sulfite.
CC   -!- SUBUNIT: Homodimer.
CC   -!- INDUCTION: By phosphate starvation.
CC   -!- MASS SPECTROMETRY: MW=36413; MW_ERR=18; METHOD=MALDI.
CC   -!- MISCELLANEOUS: Its optimum pH is between 7.25 and 7.75 and optimum
CC       temperature is 35 degrees Celsius.
CC   -!- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID
CC       DEHYDROGENASES FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF061070; AAC71709.1; -.
DR   HSSP; P36234; 1GDH.
DR   InterPro; IPR002162; D_2hydac_dh.
DR   Pfam; PF00389; 2-Hacid_DH; 1.
DR   Pfam; PF02826; 2-Hacid_DH_C; 1.
DR   PROSITE; PS00065; D_2_HYDROXYACID_DH_1; FALSE_NEG.
DR   PROSITE; PS00670; D_2_HYDROXYACID_DH_2; FALSE_NEG.
DR   PROSITE; PS00671; D_2_HYDROXYACID_DH_3; FALSE_NEG.
KW   Oxidoreductase; NAD.
FT   ACT_SITE    237    237       SUBSTRATE-BINDING (BY SIMILARITY).
FT   ACT_SITE    266    266       BY SIMILARITY.
FT   ACT_SITE    292    292       BY SIMILARITY.
SQ   SEQUENCE   336 AA;  36415 MW;  7F55D246CA4454F7 CRC64;

  Query Match             44.3%;  Score 39;  DB 1;  Length 336;
  Best Local Similarity   61.5%;  Pred. No. 23;
  Matches    8;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       4 ITHRIHWESASLL 16
           ||||:| |   ||
Db       7 ITHRVHDEILQLL 19


RESULT 11
CYDC_BACSU
ID   CYDC_BACSU      STANDARD;      PRT;   567 AA.
AC   P94366;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Transport ATP-binding protein cydC.
GN   CYDC.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / BGSC1A1;
RX   MEDLINE=97124196; PubMed=8969509;
RA   Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA   Miwa Y., Fujita Y.;
RT   "Sequencing of a 65 kb region of the Bacillus subtilis genome
RT   containing the lic and cel loci, and creation of a 177 kb contig
RT   covering the gnt-sacXY region.";
RL   Microbiology 142:3113-3123(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: SOMEHOW INVOLVED IN THE CYTOCHROME D BRANCH OF AEROBIC
CC       RESPIRATION. SEEMS TO BE A COMPONENT OF A TRANSPORT SYSTEM (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MSBA SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D83026; BAA11729.1; -.
DR   EMBL; Z99123; CAB15900.1; -.
DR   SubtiList; BG11927; cydC.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001140; ABCtranprtrTM.
DR   Pfam; PF00005; ABC_tran; 1.
```

```
DR   Pfam; PF00664; ABC_membrane; 1.
DR   ProDom; PD000006; ABC_transportr; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   ATP-binding; Transport; Transmembrane; Complete proteome.
FT   TRANSMEM     14     34       POTENTIAL.
FT   TRANSMEM     44     64       POTENTIAL.
FT   TRANSMEM    130    150       POTENTIAL.
FT   TRANSMEM    156    176       POTENTIAL.
FT   TRANSMEM    240    260       POTENTIAL.
FT   TRANSMEM    266    286       POTENTIAL.
FT   NP_BIND     360    367       ATP (POTENTIAL).
SQ   SEQUENCE   567 AA;  62806 MW;  74F2500E08C6637D CRC64;

  Query Match             44.3%;  Score 39;  DB 1;  Length 567;
  Best Local Similarity   83.3%;  Pred. No. 41;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 THRIHW 10
           |||:||
Db     519 THRLHW 524


RESULT 12
RECA_MYCTU
ID   RECA_MYCTU     STANDARD;      PRT;   790 AA.
AC   P26345; O34519;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   RecA protein (Recombinase A) [Contains: Endonuclease PI-MtuI
DE   (EC 3.1.-.-) (Mtu recA intein)].
GN   RECA OR RV2737C OR MT2806 OR MTV002.02C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=91358354; PubMed=1909321;
RA   Davis E.O., Sedgwick S.G., Colston M.J.;
RT   "Novel structure of the recA locus of Mycobacterium tuberculosis
RT   implies processing of the gene product.";
RL   J. Bacteriol. 173:5653-5662(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Canetti, and SO93;
RA   Vansoolingen D., Hoogenboezem T., Dehaas P.E., Hermans P.W.M.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   PROTEIN SPLICING.
RX   MEDLINE=93046621; PubMed=1423588;
RA   Davis E.O., Jenner P.J., Brooks P.C., Colston M.J., Sedgwick S.G.;
RT   "Protein splicing in the maturation of M. tuberculosis recA protein:
RT   a mechanism for tolerating a novel class of intervening sequence.";
RL   Cell 71:201-210(1992).
RN   [6]
RP   CHARACTERIZATION.
RX   MEDLINE=96229901; PubMed=8639660;
RA   Kumar R.A., Vaze M.B., Chandra N.R., Vijayan M., Muniyappa K.;
RT   "Functional characterization of the precursor and spliced forms of
RT   RecA protein of Mycobacterium tuberculosis.";
RL   Biochemistry 35:1793-1802(1996).
RN   [7]
RP   REVIEW.
RA   Colston M.J., Davis E.O.;
RL   (In) Bloom B.R. (eds.);
RL   Tuberculosis: pathogenesis, protection and control, pp.217-226,
RL   American Society for Microbiology, Washington DC (1994).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX   MEDLINE=20572535; PubMed=11121488;
RA   Datta S., Prabu M.M., Vaze M.B., Ganesh N., Chandra N.R.,
RA   Muniyappa K., Vijayan M.;
RT   "Crystal structures of Mycobacterium tuberculosis RecA and its
RT   complex with ADP-AlF(4): implications for decreased ATPase activity
RT   and molecular aggregation.";
RL   Nucleic Acids Res. 28:4964-4973(2000).
CC   -!- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF
CC       SINGLE-STRANDED DNA, THE ATP-DEPENDENT UPTAKE OF SINGLE-STRANDED
CC       DNA BY DUPLEX DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF
CC       HOMOLOGOUS SINGLE-STRANDED DNAS. IT INTERACTS WITH LEXA CAUSING
CC       ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE.
CC   -!- FUNCTION: PI-MTUI IS AN ENDONUCLEASE.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
CC       A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
CC       FOLLOWED BY PEPTIDE LIGATION.
CC   -!- SIMILARITY: BELONGS TO THE RECA FAMILY.
CC   -!- SIMILARITY: IN THE INTEIN SECTION; BELONGS TO THE HOMING
CC       ENDONUCLEASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X58485; CAA41395.1; -.
DR   EMBL; AL008967; CAA15533.1; -.
DR   EMBL; AJ000012; CAA03857.1; -.
DR   EMBL; AJ000011; CAA03856.1; -.
DR   EMBL; AE007109; AAK47127.1; -.
DR   PIR; S18206; S18206.
DR   PDB; 1G18; 03-JAN-01.
DR   PDB; 1G19; 03-JAN-01.
DR   REBASE; 2629; PI-MtuI.
DR   TIGR; MT2806; -.
DR   TubercuList; Rv2737c; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003586; Hedgehog_hintC.
DR   InterPro; IPR003587; Hedgehog_hintN.
DR   InterPro; IPR002203; Intein.
DR   InterPro; IPR004042; Intein_endonuc.
DR   InterPro; IPR001553; RecA.
DR   Pfam; PF00154; recA; 2.
DR   PRINTS; PR00379; INTEIN.
```

```
DR   PRINTS; PR00142; RECA.
DR   ProDom; PD000229; RecA; 2.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00305; HintC; 1.
DR   SMART; SM00306; HintN; 1.
DR   PROSITE; PS50818; INTEIN_C_TER; 1.
DR   PROSITE; PS50819; INTEIN_ENDONUCLEASE; 1.
DR   PROSITE; PS50817; INTEIN_N_TER; 1.
DR   PROSITE; PS00321; RECA_1; 1.
DR   PROSITE; PS50162; RECA_2; 1.
DR   PROSITE; PS50163; RECA_3; 2.
KW   DNA damage; DNA recombination; SOS response; ATP-binding; DNA-binding;
KW   Autocatalytic cleavage; Protein splicing; Hydrolase; Nuclease;
KW   Endonuclease; Intron homing; Complete proteome; 3D-structure.
FT   CHAIN         1    251       RECA PROTEIN, 1ST PART.
FT   CHAIN       252    691       ENDONUCLEASE PI-MTUI.
FT   CHAIN       692    790       RECA PROTEIN, 2ND PART.
FT   NP_BIND      67     74       ATP (BY SIMILARITY).
FT   VARIANT     305    305       R -> Q (IN STRAINS CANETTI AND SO93).
FT   VARIANT     430    430       A -> L (IN STRAINS CANETTI AND SO93).
FT   VARIANT     434    435       QQ -> RR (IN STRAINS CANETTI AND SO93).
FT   VARIANT     438    439       IY -> VH (IN STRAINS CANETTI AND SO93).
SQ   SEQUENCE   790 AA;  85389 MW;  AD16340D2ED5E572 CRC64;

  Query Match             44.3%;  Score 39;  DB 1;  Length 790;
  Best Local Similarity   50.0%;  Pred. No. 58;
  Matches    5;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SSKITHRIHW 10
          | :: |:|||
Db    492 SEQLAHQIHW 501


RESULT 13
OBP_HSV11
ID   OBP_HSV11      STANDARD;      PRT;   851 AA.
AC   P10193;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Origin of replication binding protein.
GN   UL9.
OS   Herpes simplex virus (type 1 / strain 17).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88274327; PubMed=2839594;
RA   McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA   McNab D., Perry L.J., Scott J.E., Taylor P.;
RT   "The complete DNA sequence of the long unique region in the genome of
RT   herpes simplex virus type 1.";
RL   J. Gen. Virol. 69:1531-1574(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88091053; PubMed=2826807;
RA   McGeoch D.J., Dalrymple M.A., Dolan A., McNab D., Perry L.J.,
RA   Taylor P., Challberg M.D.;
RT   "Structures of herpes simplex virus type 1 genes required for
RT   replication of virus DNA.";
RL   J. Virol. 62:444-453(1988).
CC   -!- FUNCTION: PROBABLY INVOLVED IN DNA REPLICATION. BINDS THE ORIGIN
CC       OF REPLICATION (ORI).
CC   -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL9,
CC       EHV-1 53, AND VZV 51.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10879; BAA01655.1; -.
DR   EMBL; X14112; CAA32345.1; -.
DR   EMBL; M19120; AAA45822.1; -.
DR   PIR; B29890; WMBEU9.
DR   PIR; I28133; I28133.
DR   TRANSFAC; T00957; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR003450; Herpes_ori_bp.
DR   Pfam; PF02399; Herpes_ori_bp; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   DNA replication; DNA-binding; ATP-binding.
FT   NP_BIND      81     88       ATP (POTENTIAL).
SQ   SEQUENCE   851 AA;  94261 MW;  961A133FE7A30CA7 CRC64;

  Query Match             44.3%;  Score 39;  DB 1;  Length 851;
  Best Local Similarity   46.2%;  Pred. No. 63;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 SKITHRIHWESAS 14
          | :  |:||:||:
Db    648 STMAARLHWDSAA 660


RESULT 14
TNP3_ECOLI
ID   TNP3_ECOLI     STANDARD;      PRT;  1015 AA.
AC   P03008;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Transposase for transposon Tn3.
GN   TNPA.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TRANSPOSON=Tn3;
RX   MEDLINE=80090058; PubMed=391406;
RA   Heffron F., McCarthy B.J., Ohtsubo H., Ohtsubo E.;
RT   "DNA sequence analysis of the transposon Tn3: three genes and three
RT   sites involved in transposition of Tn3.";
RL   Cell 18:1153-1163(1979).
CC   -!- FUNCTION: REQUIRED FOR TRANSPOSITION OF TRANSPOSON TN3.
CC   -!- SIMILARITY: BELONGS TO THE TRANSPOSASE FAMILY 7.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; V00613; CAA23884.1; ALT_SEQ.
DR   PIR; A03538; TQECT.
DR   InterPro; IPR002513; Transposase_7.
DR   Pfam; PF01526; Transposase_7; 1.
KW   Transposable element; Transposition; DNA-binding; DNA recombination.
SQ   SEQUENCE   1015 AA;  114529 MW;  571A442203B5FACA CRC64;

  Query Match             44.3%;  Score 39;  DB 1;  Length 1015;
  Best Local Similarity   50.0%;  Pred. No. 76;
  Matches    6;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      6 HRIHWESASLLR 17
          ||::|  |: ||
Db    656 HRLNWTKANYLR 667
```

```
RESULT 15
DSCA_HUMAN
ID   DSCA_HUMAN     STANDARD;      PRT;  2012 AA.
AC   O60469; O60468;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Down syndrome cell adhesion molecule precursor (CHD2).
GN   DSCAM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Brain;
RX   MEDLINE=98087574; PubMed=9426258;
RA   Yamakawa K., Huot Y.-K., Haendelt M.A., Hubert R., Chen X.-N.,
RA   Lyons G.E., Korenberg J.R.;
RT   "DSCAM: a novel member of the immunoglobulin superfamily maps in a
RT   Down syndrome region and is involved in the development of the
RT   nervous system.";
RL   Hum. Mol. Genet. 7:227-237(1998).
RN   [2]
RP   SEQUENCE FROM N.A., AND FUNCTION.
RX   MEDLINE=20384934; PubMed=10925149;
RA   Agarwala K.L., Nakamura S., Tsutsumi Y., Yamakawa K.;
RT   "Down syndrome cell adhesion molecule DSCAM mediates homophilic
RT   intercellular adhesion.";
RL   Brain Res. Mol. Brain Res. 79:118-126(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20289799; PubMed=10830953;
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA   Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA   Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA   Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA   Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RT   "The DNA sequence of human chromosome 21.";
RL   Nature 405:311-319(2000).
CC   -!- FUNCTION: CELL ADHESION MOLECULE THAT CAN MEDIATE CATION-
CC       INDEPENDENT HOMOPHILIC BINDING ACTIVITY. COULD BE INVOLVED IN
CC       NERVOUS SYSTEM DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE). THE
CC       SHORT ISOFORM MAY BE SECRETED.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM/CHD2-52 (SHOWN HERE)
CC       AND A SHORT FORM/CHD2-42; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 10 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF023450; AAC17967.1; -.
DR   EMBL; AF023449; AAC17966.1; -.
DR   EMBL; AF217525; AAF27525.1; -.
DR   EMBL; AL163283; CAB90464.1; -.
DR   EMBL; AL163282; CAB90436.1; -.
DR   EMBL; AL163281; CAB90444.1; -.
DR   Genew; HGNC:3039; DSCAM.
DR   MIM; 602523; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 10.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 7.
KW   Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat;
KW   Transmembrane; Alternative splicing.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   2012       DOWN SYNDROME CELL ADHESION MOLECULE.
FT   DOMAIN       18   1595       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1596   1616       POTENTIAL.
FT   DOMAIN     1617   2012       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       39    109       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      138    204       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      239    300       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      328    392       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      421    491       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      518    582       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN      610    676       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN      704    773       IG-LIKE C2-TYPE DOMAIN 8.
FT   DOMAIN      802    872       IG-LIKE C2-TYPE DOMAIN 9.
FT   DOMAIN      885    972       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      984   1076       FIBRONECTIN TYPE-III 2.
FT   DOMAIN     1088   1177       FIBRONECTIN TYPE-III 3.
FT   DOMAIN     1189   1273       FIBRONECTIN TYPE-III 4.
FT   DOMAIN     1300   1366       IG-LIKE C2-TYPE DOMAIN 10.
FT   DOMAIN     1380   1463       FIBRONECTIN TYPE-III 5.
FT   DOMAIN     1477   1562       FIBRONECTIN TYPE-III 6.
FT   DISULFID     46    102       BY SIMILARITY.
FT   DISULFID    145    197       BY SIMILARITY.
FT   DISULFID    246    293       BY SIMILARITY.
FT   DISULFID    335    385       BY SIMILARITY.
FT   DISULFID    428    484       BY SIMILARITY.
FT   DISULFID    525    575       BY SIMILARITY.
FT   DISULFID    617    669       BY SIMILARITY.
FT   DISULFID    711    766       BY SIMILARITY.
FT   DISULFID    809    865       BY SIMILARITY.
FT   DISULFID   1307   1359       BY SIMILARITY.
FT   CARBOHYD     28     28       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     78     78       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    470    470       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    487    487       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    512    512       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    658    658       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    666    666       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    710    710       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    748    748       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    795    795       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    924    924       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1142   1142       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1160   1160       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1250   1250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1271   1271       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1341   1341       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1488   1488       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1562   1571       NFATLNYDGS -> KEAARCKEFS (IN SHORT
FT                                ISOFORM).
FT   VARSPLIC   1572   2012       MISSING (IN SHORT ISOFORM).
FT   CONFLICT   1893   2012       HRPGDLIHLPPYLRMDFLLNRGGPGTSRDLSLGQACLEPQK
FT                                SRTLKRPTVLEPIPMEAASSASSTREGQSWQPGAVATLPQR
FT                                EGAELGQAAKMSSSQESLLDSRGHLKGNNPYAKSYTLV ->
FT                                IGQVTSYICLHTLEWTFC (IN REF. 1).
SQ   SEQUENCE   2012 AA;  222259 MW;  0E33CFB781A08334 CRC64;
```

Query Match 44.3%; Score 39; DB 1; Length 2012;
Best Local Similarity 42.9%; Pred. No. 1.6e+02;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```
Qy       1 SSKITHRIHWESAS 14
           |  :|| :|::| |
Db    1699 SLTVTHTVHYQSVS 1712
```

Search completed: February 24, 2003, 15:33:05
Job time : 12 secs